Optimisation Non-linéaire

(Pour Master Big Data & Aide à la Décision)

Présenté par:
A. Laghrib
laghrib.amine@gmail.com

2023

ENSA Khouribga, Université Sultan Moulay Slimane

Contents

A quoi serve l'optimisation ?

Optimisation et Science de données

L'optimisation interviennent aujourd'hui dans tous les secteurs d'activité d'un ingénieur de données ou d'un mathématicien, ou bien informaticien qui les utilise aussi bien

- En intelligence artificielle, deep learning
- En data Mining, Big data, exploration et représentation de données,
- En moteurs de recherche, Web crawlers,
- En génération de nombres aléatoires, de processus aléatoires (graphisme, jeux vidéos, ...)
- En optimisation de l'affichage d'un contenu en fonction des utilisateurs (aléatoires),
- En décision,
- En traitement de l'image,

A quoi nous servent l'optimisation non-linéaire ?

Elle est toujours utilisée notamment :

- Par les grandes sociétés dans leur production
- Dans le domaine de la couverture de risques (assurances)
- Gestion des serveurs où arrivent des requêtes
- En physique générale
- Ingénierie dans tout ses états
- En statistique et finance

Formulation Mathématique

Notations

Mathématiquement parlant, l'optimisation consiste à minimiser ou maximiser une fonction tout en respectant des contraintes spécifiques sur ses variables. Cela peut être exprimé à l'aide de la notation suivante :

- "x" représente le vecteur de variables, également appelées inconnues ou paramètres.
- "f" désigne la fonction objectif, une fonction scalaire de "x" que l'on souhaite soit maximiser, soit minimiser.
- "c_i" fait référence aux fonctions de contrainte, qui sont des fonctions scalaires de "x" et qui définissent diverses équations et inégalités. Ces contraintes précisent les conditions que le vecteur d'inconnues "x" doit satisfaire.

Formulation Mathématique

Problème général

En prenant compte cette notation, le problème d'optimisation non-linéaire est donné sous la forme suivante :

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

sous contraintes

$$c_i(x) \leq 0, \quad i \in E,$$

$$c_i(x) \geq 0, \quad i \in I.$$

Formulation Mathématique

Problème général

Comme exemple simple, considérons le problème d'optimisation suivant :

$$\text{minimiser } (x_1 - 2)^2 + (x_2 - 1)^2 \text{ sous contraintes}$$

$$x_1^2 - x_2 \leq 0,$$

$$x_1 + x_2 \leq 2.$$

Dans ce cas, nous cherchons à trouver les valeurs de x_1 et x_2 qui minimisent la fonction quadratique $(x_1 - 2)^2 + (x_2 - 1)^2$, tout en respectant les contraintes spécifiées.

Il s'agit d'un exemple basique de problème d'optimisation qui illustre comment trouver des solutions qui respectent des conditions spécifiques tout en optimisant un objectif défini.

Exemple

Formulation

Nous pouvons formuler ce problème sous la forme générale en définissant :

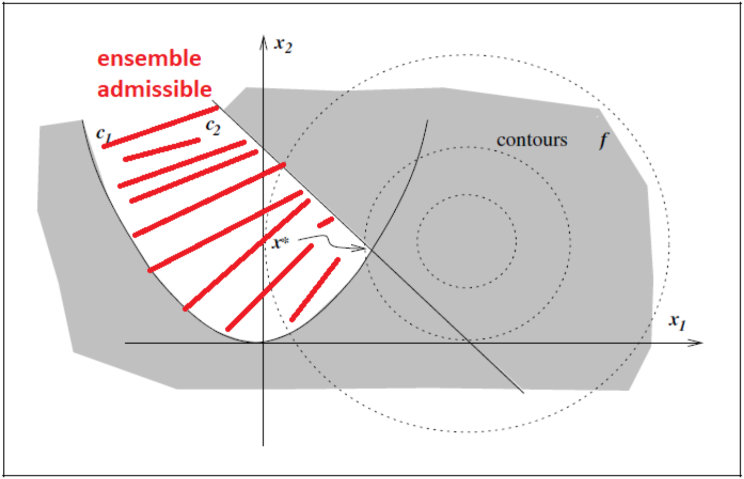
$$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

$$c(x) = \begin{bmatrix} -x_1^2 + x_2 \\ -x_1 - x_2 + 2 \end{bmatrix}, \quad I = \{1, 2\}, \quad E = \emptyset.$$

Le vecteur x contient les variables x_1 et x_2. La fonction de contrainte $c(x)$ est un vecteur qui exprime les relations d'inégalité définies dans les contraintes du problème. I représente l'ensemble des indices des contraintes, et E est l'ensemble vide car il n'y a pas de contraintes d'égalité dans cet exemple.

Exemple

Représentation graphique



Exemple

Représentation graphique

La Figure 1.1 montre les contours de la fonction objective, c'est-à-dire l'ensemble des points pour lesquels $f(x)$ a une valeur constante. Elle illustre également la région réalisable, qui est l'ensemble des points satisfaisant toutes les contraintes (la zone en rouge)

Optimisation non-linéaire

La nature du problème

Les problèmes ayant la forme générale (1.1) peuvent être classés en fonction de :

- La nature de la fonction objective et des contraintes (linéaires, non linéaires, convexes),
- du nombre de variables (grand ou petit),
- de la régularité des fonctions (différentiables ou non différentiables), etc.

Une distinction importante est celle entre les problèmes qui comportent des contraintes sur les variables et ceux qui n'en comportent pas. Ce cours est divisé en deux parties selon cette classification.

Outils Mathématiques

L'espace $\mathbb{R}^n$:

- $\mathbb{R}^n$: l'ensemble des vecteurs colonnes de n dimensions avec des composants réels, doté de l'opérateur d'addition composant par composant :

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Outils Mathématiques

L'espace $\mathbb{R}^n$:

- Le produit d'un scalaire λ par un vecteur :

$$\lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

- $e_1, e_2, \ldots, e_n$: la base standard/canonique.
- e et 0 : les vecteurs colonnes de tous "1" (vecteur unité) et de tous "0" (vecteur nul).

Outils Mathématiques

Sous espace de $\mathbb{R}^n$:

- Orthant non négatif :

$$\mathbb{R}^n_+ = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : x_1, x_2, \ldots, x_n \geq 0 \right\}$$

- Orthant positif :

$$\mathbb{R}^n_{++} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : x_1, x_2, \ldots, x_n > 0 \right\}$$

Outils Mathématiques

Les segments de $\mathbb{R}^n$:

Les segments de droite fermé $[x, y]$ et ouvert (x, y) en $\mathbb{R}^n$, ainsi que leur définition :

- Segment de droite fermé :

$$[x, y] = \{x + \alpha(y - x) : \alpha \in [0, 1]\}$$

- Segment de droite ouvert :

$$(x, y) = \{x + \alpha(y - x) : \alpha \in (0, 1)\}$$

Ces ensembles sont définis pour $x \neq y$, et $(x, x) = \emptyset$.

Outils Mathématiques

Les segments de $\mathbb{R}^n$:

Lorsque $x \neq y$, $\Delta_n = 0$. Le simplexe unitaire, noté Δ_n, est un sous-ensemble de $\mathbb{R}^n$ comprenant tous les vecteurs non négatifs dont la somme est égale à 1 :

$$\Delta_n = \{x \in \mathbb{R}^n : x > 0, \mathbf{e}^T x = 1\}.$$

Outils Mathématiques

Ensemble des matrices réelles

Ensemble des matrices réelles :

$$\mathbb{R}^{m \times n}$$

- Matrice identité de $n \times n$:

$$\mathbf{I}_n$$

- Matrice de zéros de $m \times n$:

$$\mathbf{0}_{m \times n}$$

Outils Mathématiques

Produit scalaire

Définition Un produit scalaire sur $\mathbb{R}^n$ est une fonction $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ vérifiant les propriétés suivantes :

1. (*symétrie*) $\langle x, y \rangle = \langle y, x \rangle$ pour tout $x, y \in \mathbb{R}^n$.
2. (*additivité*) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ pour tout $x, y, z \in \mathbb{R}^n$.
3. (*homogénéité*) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ pour tout $\lambda \in \mathbb{R}$ et $x, y \in \mathbb{R}^n$.
4. (*positivité définitive*) $\langle x, x \rangle \geq 0$ pour tout $x \in \mathbb{R}^n$, et $\langle x, x \rangle = 0$ si et seulement si $x = \mathbf{0}$.

Outils Mathématiques

Produit scalaire

Exemples

- Le *produit scalaire usuel* ou *produit scalaire canonique*, noté $\langle x, y\rangle = x^T y$, est donné par :

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i \quad \text{pour tout } x, y \in \mathbb{R}^n.$$

- Le *produit scalaire pondéré*, noté $\langle x, y\rangle_w$, est donné par :

$$\langle x, y\rangle_w = \sum_{i=1}^{n} w_i x_i y_i, \qquad \text{où } w \in \mathbb{R}^{n++}.$$

Outils Mathématiques

Norme sur $\mathbb{R}^n$

Définition. Une norme $\|\cdot\|$ sur $\mathbb{R}^n$ est une fonction $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ qui satisfait les propriétés suivantes :

- *(Non-négativité)* $\|x\| \geq 0$ pour tout $x \in \mathbb{R}^n$ et $\|x\| = 0$ si et seulement si $x = 0$.
- *(Homogénéité positive)* $\|\lambda x\| = |\lambda|\|x\|$ pour tout $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$.
- *(Inégalité triangulaire)* $\|x + y\| \leq \|x\| + \|y\|$ pour tout $x, y \in \mathbb{R}^n$.

Outils Mathématiques

Norme sur $\mathbb{R}^n$

Une manière naturelle de définir une norme sur $\mathbb{R}^n$ est de prendre un produit interne $\langle \cdot, \cdot \rangle$ défini sur $\mathbb{R}^n$, et de définir la norme associée comme $\|x\| \equiv \sqrt{\langle x, x \rangle}$ pour tout $x \in \mathbb{R}^n$.

- La norme associée au produit scalaire est la norme euclidienne ou l_2-norme :

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \quad \text{pour tout } x \in \mathbb{R}^n$$

Outils Mathématiques

l^p Normes

La norme l_p $(p \geq 1)$ est définie par $\|x\|_p \equiv \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$.

- La norme l_∞ est définie comme : $\|x\|_\infty \equiv \max_{i=1,2,\ldots,n} |x_i|$.
- On peut montrer que $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$.

Outils Mathématiques

$\natural^p$ Normes

Exemple : $l_{\frac{1}{2}}$ n'est pas une norme. Pourquoi ?

Outils Mathématiques

l^p Normes

Exemple : $l_{\frac{1}{2}}$ n'est pas une norme. Pourquoi ?

Réponse

La norme $l_{\frac{1}{2}}$ est définie par $\|x\|_{\frac{1}{2}} = \left(\sum_{i=1}^{n} |x_i|^{\frac{1}{2}}\right)^2$, ce qui ne satisfait pas l'inégalité triangulaire ($\|x+y\|_{\frac{1}{2}} \leq \|x\|_{\frac{1}{2}} + \|y\|_{\frac{1}{2}}$) pour tous $x, y \in \mathbb{R}^n$. Ainsi, $l_{\frac{1}{2}}$ ne constitue pas une norme.

Outils Mathématiques

Cauchy-Schwarz

L'inégalité de Cauchy-Schwarz affirme que, pour tout x et y dans un espace vectoriel muni d'un produit scalaire, la valeur absolue du produit scalaire entre x et y est inférieure ou égale au produit des normes de x et y :

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

C'est une inégalité fondamentale avec de nombreuses applications en mathématiques et dans divers domaines appliqués.

Outils Mathématiques

Normes matricielles

Définition. Une norme $\|\,.\,\|$ sur $\mathbb{R}^{m\times n}$ est une fonction

$$\|\,.\,\| : \mathbb{R}^{m\times n} \to \mathbb{R}$$

qui satisfait les propriétés suivantes :
1. (Non-négativité) $\|A\| \geq 0$ pour tout $A \in \mathbb{R}^{m\times n}$, et $\|A\| = 0$ si et seulement si $A = 0$.
2. (Homogénéité positive) $\|\lambda A\| = |\lambda|\|A\|$ pour tout $A \in \mathbb{R}^{m\times n}$ et $\lambda \in \mathbb{R}$.
3. (Inégalité triangulaire) $\|A + B\| \leq \|A\| + \|B\|$ pour tout $A, B \in \mathbb{R}^{m\times n}$.

Outils Mathématiques

Normes matricielles

Norme spectrale : La norme induite (2,2) d'une matrice $A \in \mathbb{R}^{m \times n}$ est la plus grande valeur singulière de A :

$$\|A\|_2 = \|A\|_{2,2} = \sqrt{\lambda_{\max}(A^T A)} \equiv \sigma_{\max}(A)$$

Cette norme est appelée norme spectrale.

Norme l_1 : Lorsque $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, la norme matricielle (1,1) induite d'une matrice $A \in \mathbb{R}^{m \times n}$ est donnée par :

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |A_{i,j}|$$

Outils Mathématiques

Normes matricielles

Norme l_∞ : Lorsque $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, la norme matricielle (∞, ∞) induite d'une matrice $A \in \mathbb{R}^{m\times n}$ est donnée par :

$$\|A\|_\infty = \max_{i=1,2,\ldots,m} \sum_{j=1}^{n} |A_{i,j}|$$

Outils Mathématiques

valeur et vecteur propre

- Soit $A \in \mathbb{R}^{n \times n}$. Alors, un vecteur non nul $v \in \mathbb{R}^n$ est appelé un vecteur propre de A s'il existe un $\lambda \in \mathbb{C}$ tel que

$$Av = \lambda v$$

- Le scalaire λ est la valeur propre correspondant au vecteur propre v.
- En général, les matrices à valeurs réelles peuvent avoir des valeurs propres complexes, mais lorsque la matrice est symétrique, les valeurs propres sont nécessairement réelles.

Outils Mathématiques

Exemple

Supposons que nous ayons la matrice suivante A :

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

Outils Mathématiques

Exemple

Supposons que nous ayons la matrice suivante A :

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

Pour trouver les valeurs propres de A, nous devons résoudre l'équation caractéristique $\det(A - \lambda I) = 0$, où λ est la valeur propre que nous cherchons, et I est la matrice identité.

Outils Mathématiques

Exemple

L'équation caractéristique est :

$$\det\left(\begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}\right) = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = 0$$

Nous résolvons cette équation quadratique pour λ en utilisant la formule quadratique :

$$\lambda = \frac{-(-6) \pm \sqrt{(-6)^2 - 4 \times 1 \times 8}}{2 \times 1}$$

Outils Mathématiques

Exemple

Simplifiant, nous obtenons :

$$\lambda_1 = 4$$

$$\lambda_2 = 2$$

Donc, les valeurs propres de la matrice A sont $\lambda_1 = 4$ et $\lambda_2 = 2$.

Outils Mathématiques

valeur et vecteur propre

- Les valeurs propres d'une matrice symétrique $n \times n$ A sont notées comme suit :

$$\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$$

- La plus grande valeur propre est également notée $\lambda_{\text{max}}(A)$ $(= \lambda_1(A))$,
- La plus petite valeur propre est également notée $\lambda_{\text{min}}(A)$ $(= \lambda_n(A))$.

Outils Mathématiques

Décomposition

Théorème : Soit $A \in \mathbb{R}^{n\times n}$ une matrice symétrique $n \times n$. Alors, il existe une matrice orthogonale $U \in \mathbb{R}^{n\times n}$ ($U^T U = UU^T = I$) et une matrice diagonale $D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ telles que :

$$U^T A U = D.$$

-

Les colonnes de la matrice U constituent une base orthogonale comprenant les vecteurs propres de A, et les éléments diagonaux de D sont les valeurs propres correspondantes.

- Un résultat direct est que $\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i(A)$ et $\det(A) = \prod_{i=1}^{n} \lambda_i(A)$.

Outils Mathématiques

Concepts topologiques de base :

- La boule ouverte de centre c dans $\mathbb{R}^n$ et de rayon r est définie par :

$$B(c,r) = \{x : \|x - c\| < r\}$$

- La boule fermée de centre c et de rayon r est définie par :

$$B[c,r] = \{x : \|x - c\| \leq r\}$$

Outils Mathématiques

Définition

Étant donné un ensemble U contenu dans $\mathbb{R}^n$, un point c appartenant à U est appelé un point intérieur de U s'il existe $r > 0$ tel que $B(c, r) \subseteq U$. L'ensemble de tous les points intérieurs d'un ensemble donné U est appelé l'intérieur de l'ensemble et est noté $\text{int}(U)$:

$$\text{int}(U) = \{x \in U : B(x, r) \subseteq U \text{ pour un certain } r > 0\}$$

Outils Mathématiques

Ensemble compacte

Définitions :

- Un ensemble U contenu dans $\mathbb{R}^n$ est appelé borné s'il existe un $M > 0$ tel que U soit contenu dans $B(0, M)$.
- Un ensemble U contenu dans $\mathbb{R}^n$ est appelé compact s'il est fermé et borné.

Exemples d'ensembles compacts : boules fermées, simplexe unité, segments de droite fermés.

Outils Mathématiques

Dérivées directionnelles

Définitions :

- Soit f une fonction définie sur un ensemble S contenu dans $\mathbb{R}^n$. Soit $x \in \text{int}(S)$ et $d \in \mathbb{R}^n$. Si la limite

$$\lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t}$$

existe, alors elle est appelée la dérivée directionnelle de f en x dans la direction d et est notée $f'(x; d)$.

- Pour tout $i = 1, 2, \ldots, n$, si la limite

$$\lim_{t \to 0} \frac{f(x+te_i) - f(x)}{t}$$

existe, alors sa valeur est appelée la dérivée partielle par rapport à x_i et est notée $\frac{\partial f}{\partial x_i}(x)$.

Outils Mathématiques

Dérivées directionnelles

- Si toutes les dérivées partielles d'une fonction f existent en un point $x \in \mathbb{R}^n$, alors le gradient de f en x est

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

Outils Mathématiques

Differentiabilité au sens de Gâteaux

Une fonction f définie sur un ensemble ouvert U de $\mathbb{R}^n$ est appelée continûment différentiable sur U si toutes les dérivées partielles existent et sont continues sur U. Dans ce cas, $f'(x;d) = \nabla f(x)^T d$, $x \in U, d \in \mathbb{R}^n$.

Proposition

Soit $f : U \to \mathbb{R}$ définie sur un ensemble ouvert $U \subseteq \mathbb{R}^n$. Supposons que f soit continûment différentiable sur U. Alors,

$$\lim_{d \to 0} \frac{f(x+d) - f(x) - \nabla f(x)^T d}{\|d\|} = 0$$

pour tout $x \in U$.

Outils Mathématiques

Différentiabilité au sens de Gâteaux

Une autre façon d'écrire le résultat ci-dessus est la suivante :

$$f(y) = f(x) + \nabla f(x)^T (y - x) + o(\|y - x\|)$$

où $o(\cdot) : \mathbb{R}_+^n \to \mathbb{R}$ est une fonction unidimensionnelle telle que $\frac{o(t)}{t} \to 0$ lorsque $t \to 0^+$.

Outils Mathématiques

Dérivée seconde

- Les dérivées partielles $\frac{\partial f}{\partial x_i}$ sont elles-mêmes des fonctions réelles qui peuvent être partiellement dérivées. Les dérivées partielles d'ordre (i, j) de f en $x \in U$ (si elles existent) sont définies par

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

- Un fonction f définie sur un ensemble ouvert $U \subset \mathbb{R}^n$ est dite deux fois continûment différentiable sur U si toutes les dérivées partielles d'ordre deux existent et sont continues sur U. Dans ce cas, pour tout $i \neq j$ et tout $x \in U$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

Outils Mathématiques

La matrice Hessienne

Le hessien de f en un point $x \in U$ est la matrice $n \times n$ définie par :

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

Pour les fonctions deux fois continûment différentiables, le hessien est une matrice symétrique.

Outils Mathématiques

Théorème d'approximation linéaire

Soit $f : U \to \mathbb{R}$ définie sur un ensemble ouvert $U \subset \mathbb{R}^n$. Supposons que f est deux fois continûment différentiable sur U. Soit $x \in U$ et $r > 0$ tel que $B(x, r) \subset U$. Alors, pour tout $y \in B(x, r)$, il existe ξ dans le segment $[x, y]$ tel que :

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\xi)(y - x)$$

Contents

Optimisation sans contraintes

Définition (**Minimum**)

- On cherche à résoudre

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

avec $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et $\Omega \subseteq \mathbb{R}^n$

- Un point réalisable $\mathrm{x}^* \in \Omega$ est un minimum global de la fonction f sur le domaine Ω si

$$f\left(\mathbf{x}^*\right) \leq f(\mathbf{x}) \text{ pout tout } \mathbf{x} \in \Omega$$

Optimisation sans contraintes

Définition (Minimum et Maximum)

Soit $f : S \to \mathbb{R}$ définie sur un ensemble $S \subseteq \mathbb{R}^n$. Alors,

1. $x^* \in S$ est un point de minimum global de f sur S si $f(x) \geq f(x^*)$ pour tout $x \in S$.
2. $x^* \in S$ est un point de minimum global strict de f sur S si $f(x) > f(x^*)$ pour tout $x \neq x^*$ dans S.
3. $x^* \in S$ est un point de maximum globalde f sur S si $f(x) \leq f(x^*)$ pour tout $x \in S$.
4. $x^* \in S$ est un point de maximum global strict de f sur S si $f(x) < f(x^*)$ pour tout $x \neq x^*$ dans S.

Optimisation sans contraintes

Définition (**Minimum et Maximum**)

Un optimum global est soit un minimum global, soit un maximum global.

- La valeur maximale de f sur S :

$$\sup\{f(x) : x \in S\}$$

- Valeur minimale de f sur S :

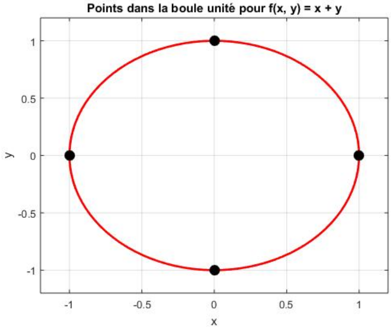
$$\inf\{f(x) : x \in S\}$$

- Les valeurs minimale et maximale sont toujours uniques.

Optimisation sans contraintes

Définition (**Exemple**)

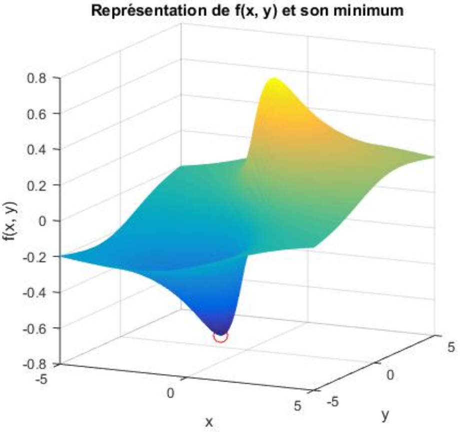
Pour la fonction $f(x, y) = x + y$ sur la boule unité $S = B[0, 1] = \{(x, y) : x^2 + y^2 \leq 1\}$, cherchons le minimum et le maximum global.



Optimisation sans contraintes

Définition (**Exemple 2**)

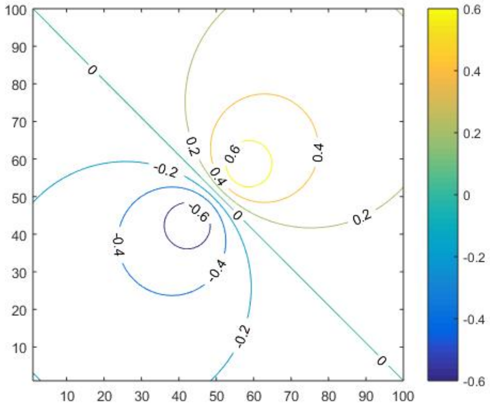
$$\min_{x,y}\left(f(x,y)=\frac{x+y}{x^2+y^2+1}\right)$$



Optimisation sans contraintes

Définition (**Exemple 2**)

Coordonnées du minimum : $x = -0.70711$, $y = -0.70711$. La valeur minimale de $f(x, y)$: -0.70711.



Optimisation sans contraintes

Définition (**Minimum**)

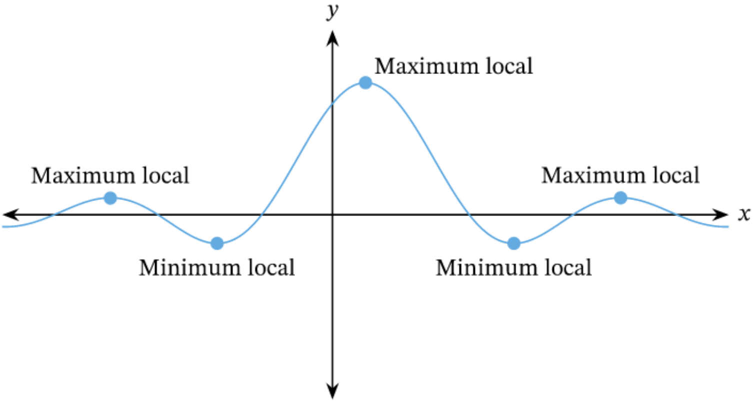
- Un point réalisable $\mathbf{x}^* \in \Omega$ est un minimum local de f sur Ω s'il existe $\varepsilon > 0$ tel que :

$$f\left(\mathbf{x}^*\right) \leq f(\mathbf{x}) \text{ pout tout } \mathbf{x} \in \Omega \cap \mathcal{B}_{\varepsilon}\left(\mathbf{x}^*\right)$$

avec $\mathcal{B}_{\varepsilon}\left(\mathbf{x}^*\right) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon\}$

Optimisation sans contraintes

Représentation graphique



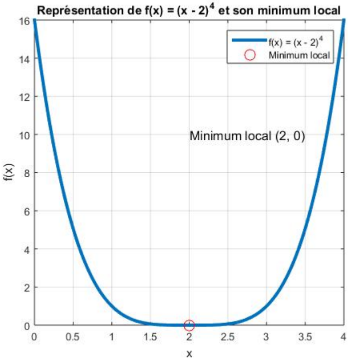
Optimisation sans contraintes

Remarque

Pour la fonction $f(x) = 2$, pour chaque point x est minimum local, tandis que la fonction $f(x) = (x-2)^4$ possède un seul strict minimum local dans $x = 2$.

Optimisation sans contraintes

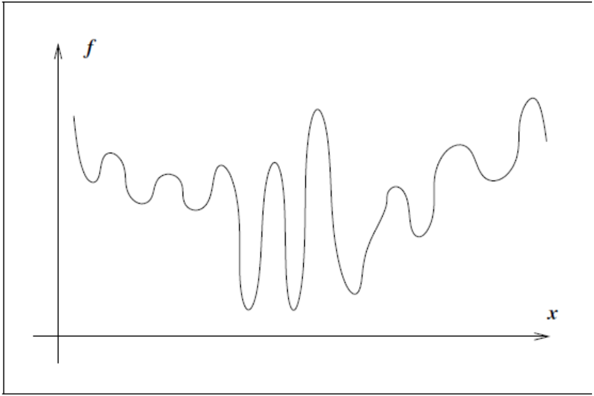
Représentation graphique



Optimisation sans contraintes

Illustration

Une fonction avec de nombreux minimiseurs locaux.



Optimisation sans contraintes

Quelques Notions

- Gradient de f en $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T \in \mathbb{R}^n$$

- Dérivée directionnelle de f en $\mathbf{x} \in \mathbb{R}^n$ dans la direction unitaire $\mathbf{d} \in \mathbb{R}^n$:

$$f'_{\mathbf{d}}(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} = \mathbf{d}^\top \nabla f(\mathbf{x})$$

Optimisation sans contraintes

Quelques Notions

- Si les dérivées secondes de f existent et sont continues, alors la matrice hessienne en x s'écrit

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

Optimisation sans contraintes

Direction de descente

- $\mathbf{d} \in \mathbb{R}^n$ est une direction (stricte) de descente de f en $\mathbf{x} \in \mathbb{R}^n$ si

$$f'_{\mathbf{d}}(\mathbf{x}) = \mathbf{d}^\top \nabla f(\mathbf{x}) < 0$$

- Pour tout $\alpha \in \mathbb{R}^+$ petit, on aura $h(\alpha) = f(\mathbf{x} + \alpha\mathbf{d}) < f(\mathbf{x})$ et $h'(0) < 0$
- Principe de la line search (recherche linéaire) : Trouver α tel que $h'(\alpha) = 0$

Optimisation sans contraintes

Signe d'une matrice

$A \in \mathbb{R}^{n\times n}$ symétrique est dite

- Semi-définie positive si $\mathbf{x}^\top A\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie positive si $\mathbf{x}^\top A\mathbf{x} > 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Semi-définie négative si $\mathbf{x}^\top A\mathbf{x} \leq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie négative si $\mathbf{x}^\top A\mathbf{x} < 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Indéfinie sinon

Optimisation sans contraintes

Signe d'une matrice

- La définie positivité d'une matrice ne signifie pas que ses composants sont positifs, comme le montrent les exemples suivants.

Exemple 1

Soit $A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. Ainsi, A est semi-définie positive. En fait, puisque $x_1^2 + (x_1 - x_2)^2 = 0$ si et seulement si $x_1 = x_2 = 0$, il en résulte que A est définie positive. Cet exemple illustre le fait qu'une matrice définie positive peut avoir des composantes négatives.

Optimisation sans contraintes

Exemple 2

Soit $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.

Optimisation sans contraintes

Exemple 2

Soit $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Cette matrice, dont les composants sont tous positifs, n'est pas définie positive car pour $x = (1, -1)$, nous avons $x^T A x = (1, -1) \times \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2$. Cela démontre qu'elle ne satisfait pas aux critères de positivité définie.

Optimisation sans contraintes

Matrice diagonal

Lemme

Soit $A \in \mathbb{R}^{n\times n}$ *une matrice définie positive. Alors, les éléments de la diagonale de* A *sont positifs.*

Optimisation sans contraintes

Matrice définie positive

Definition

Soit $A \in \mathbb{R}^{n \times n}$ une matrice symétrique.

1. A est appelée négativement semi-définie, noté $A \preceq 0$, si $x^T Ax < 0$ pour tout $x \in \mathbb{R}^n$.
2. A est appelée négativement définie, noté $A \prec 0$, si $x^T Ax < 0$ pour tout $x \in \mathbb{R}^n$ avec $x \neq 0$.
3. A est appelée indéfinie s'il existe $x, y \in \mathbb{R}^n$ tels que $x^T Ax > 0$ et $y^T Ay < 0$.

Optimisation sans contraintes

Matrice définie positive

Lemma

a) *Si A est une matrice négative définie, alors les éléments diagonaux de A sont négatifs.*

b) *Si A est une matrice semi-définie positive, alors les éléments diagonaux de A sont non positifs.*

Lorsque la diagonale d'une matrice contient à la fois des éléments positifs et négatifs, alors la matrice est indéfinie. L'affirmation inverse n'est pas correcte.

Optimisation sans contraintes

Matrice définie positive

Lemma

Soit A une matrice symétrique $n \times n$. Si la diagonale de A contient à la fois des éléments positifs et négatifs, alors A est indéfinie.

Optimisation sans contraintes

Théorème de caractérisation des valeurs propres

Théorème

Soit A une matrice symétrique $n \times n$. Alors :

a) *A est définie positive si et seulement si toutes ses valeurs propres sont positives,*
b) *A est semi-définie positive si et seulement si toutes ses valeurs propres sont non négatives,*
c) *A est définie négative si et seulement si toutes ses valeurs propres sont négatives,*
d) *A est semi-définie négative si et seulement si toutes ses valeurs propres sont non positives,*
e) *A est indéfinie si et seulement si elle a au moins une valeur propre positive et au moins une valeur propre négative.*

Optimisation sans contraintes

Matrice définie positive

Corollaire

Soit A une matrice semi-définie positive (définie positive). Alors, Tr(A) et det(A) sont non négatifs (positifs).

Optimisation sans contraintes

Matrice définie positive

Lemme

Soit $D = diag(d_1, d_2, \ldots, d_n)$. *Alors*

1. D *est définie positive si et seulement si* $d_i > 0$ *pour tout* i,
2. D *est semi-définie positive si et seulement si* $d_i \geq 0$ *pour tout* i,
3. D *est définie négative si et seulement si* $d_i < 0$ *pour tout* i,
4. D *est semi-définie négative si et seulement si* $d_i \leq 0$ *pour tout* i,
5. D *est indéfinie si et seulement s'il existe* i *et* j *tels que* $d_i > 0$ *et* $d_j < 0$.

Optimisation sans contraintes

Matrice définie positive

Proposition

Soit A *une matrice* 2×2 *symétrique. Alors* A *est semi-définie positive (définie) si et seulement si* $Tr(A), \det(A) \geq 0$ *(*$Tr(A), \det(A) > 0$*).*

Optimisation sans contraintes

Matrice définie positive

Example

(Example 2.21) Considérons les matrices B et A définies comme suit :

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0.1 \end{pmatrix}, \quad A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}.$$

Optimisation sans contraintes

Matrice définie positive

Example

(Example 2.21) Considérons les matrices B et A définies comme suit :

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0.1 \end{pmatrix}, \quad A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}.$$

- La matrice A est positive définitive puisque Tr$(A) = 4 + 3 = 7 > 0$ et $\det(A) = 4 \times 3 - 1 \times 1 = 11 > 0$.
- La matrice B est de taille 3×3 donc la règle ne s'applique pas.

Optimisation sans contraintes

Principal Minors Criterion

Théorème

Soit A *une matrice symétrique* $n \times n$. *Alors* A *est définie positive si et seulement si* $D_1(A) > 0$, $D_2(A) > 0$, ..., $D_n(A) > 0$.

Remarque :

Le critère des mineurs principaux est un outil pour détecter la définie positivité d'une matrice. Il ne peut pas être utilisé pour détecter la semi-définie positivité.

Optimisation sans contraintes

Principal Minors Criterion

Exemple

Soit Aet B deux matrices définies par

$$A = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & -1 \end{pmatrix}.$$

Solution :

La matrice A est définie positive. La matrice B est non définie positive.

Optimisation sans contraintes

Matrice dominantes

Definition

Soit A une matrice symétrique $n \times n$.

1. A est appelée diagonalement dominante si pour tout $i = 1, 2, \ldots, n$,

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|.$$

2. A est appelée strictement diagonalement dominante si pour tout $i = 1, 2, \ldots, n$,

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|.$$

Optimisation sans contraintes

Matrice dominantes

Théorème

(a) *Soit A une matrice diagonalement dominante $n \times n$ symétrique dont les éléments de la diagonale sont non négatifs. Alors A est positive semi-définie.*

(b) *Soit A une matrice diagonalement dominante stricte $n \times n$ symétrique dont les éléments de la diagonale sont positifs. Alors A est positive définie.*

Optimisation sans contraintes

Conditions nécessaire d'optimalité

Dans le cas $\Omega = \mathbb{R}^n$:
Condition nécessaire d'optimalité de premier ordre
Si $\mathbf{x}^*$ est un minimum local de f sur $\mathbb{R}^n$, alors $\nabla f\left(\mathbf{x}^*\right) = \mathbf{0}$ ($\mathbf{x}^*$ est un point critique)

Démonstration

Il faut passer par la formule de Taylor :

- **Formule de Taylor avec reste intégral :** pour une fonction continûment différentiable. Si $f : \mathbb{R}^n \to \mathbb{R}$ est continûment différentiable, alors pour tout $p \in \mathbb{R}^n$, il existe un t dans l'intervalle $(0, 1)$ tel que :

$$f(x+p) = f(x) + \nabla f(x+tp)^T p$$

Optimisation sans contraintes

Formule de Taylor

- **Formule de Taylor avec reste intégral :** pour une fonction deux fois continûment différentiable. Si $f : \mathbb{R}^n \to \mathbb{R}$ est deux fois continûment différentiable, alors pour tout $p \in \mathbb{R}^n$, il existe un t dans l'intervalle $(0, 1)$ tel que :

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\, dt$$

Optimisation sans contraintes

Formule de Taylor

- **Egalité de Taylor :** pour une fonction deux fois continûment différentiable : Si $f : \mathbb{R}^n \to \mathbb{R}$ est deux fois continûment différentiable, alors pour tout $p \in \mathbb{R}^n$, il existe un t dans l'intervalle $(0, 1)$ tel que :

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

où $\nabla^2 f$ est la matrice hessienne de f.

Optimisation sans contraintes

Démonstration du theorème

C'est une démonstration par l'absurde pour montrer que si $\nabla f(x^*) \neq 0$, alors x^* n'est pas un minimum local. Voici la démarche en détail :

- **Hypothèse de départ :** Supposons par contradiction que $\nabla f(x^*) \neq 0$.
- **Définition du vecteur p**
  Définissons $p = -\nabla f(x^*)$. Alors, $p^T \nabla f(x^*) < 0$ car le produit scalaire de deux vecteurs est négatif si les vecteurs forment un angle obtus ($\nabla f(x^*)$ est l'opposé de p).
- **Continuité de ∇f près de x^* :**
  Comme ∇f est continue près de x^*, il existe un $t > 0$ tel que $p^T \nabla f(x^* + tp) < 0$ pour tout $t \in [0, T]$. Cela signifie que le produit scalaire de p avec ∇f reste négatif le long de la ligne $x^* + tp$.

Optimisation sans contraintes

Démonstration du theorème

-

- **Utilisation du théorème de Taylor :**
  En utilisant le théorème de Taylor, on sait que pour tout $\bar{t}$ dans l'intervalle $(0, T]$, on a :

$$f(x^* + \bar{t}p) = f(x^*) + \bar{t}p^T \nabla f(x^* + tp)$$

  pour un certain t dans l'intervalle $(0, \bar{t})$.

- Comme $p^T \nabla f(x^* + tp) < 0$ pour tout t, cela implique que $f(x^* + \bar{t}p) < f(x^*)$ pour tout $\bar{t}$ dans l'intervalle $(0, T]$. Ainsi, le long de la direction de p, la fonction f décroît. Cela signifie que x^* n'est pas un minimum local, ce qui contredit notre hypothèse initiale.
  Par conséquent, si $\nabla f(x^*) \neq 0$, alors x^* n'est pas un minimum local.

Optimisation sans contraintes

Exemple

Examinons en détail les calculs pour trouver les points stationnaires de $f(x,y) = \frac{x+y}{(x^2+y^2+1)^2}$ ainsi que leur nature (maximum global, minimum global).
Tout d'abord, nous devons calculer le gradient de $f(x,y)$, noté $\nabla f(x,y)$, qui est un vecteur contenant les dérivées partielles de f par rapport à x et y.

Optimisation sans contraintes

Exemple

Le gradient de $f(x, y)$ est donné par :

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

Où les dérivées partielles sont calculées comme suit :

$$\frac{\partial f}{\partial x}(x, y) = \frac{-(x^2 + y^2 + 1)^2 + 2x(x + y)}{(x^2 + y^2 + 1)^4}$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{-(x^2 + y^2 + 1)^2 + 2y(x + y)}{(x^2 + y^2 + 1)^4}$$

Optimisation sans contraintes

Exemple

En simplifiant, nous obtenons :

$$\frac{\partial f}{\partial x}(x, y) = \frac{1 - 2x^2}{(x^2 + y^2 + 1)^3}$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{1 - 2y^2}{(x^2 + y^2 + 1)^3}$$

Ensuite, nous cherchons les points stationnaires en résolvant les équations suivantes pour x et y :

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

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

Optimisation sans contraintes

Exemple

Cela nous donne les équations :

$$1 - 2x^2 = 0$$

$$1 - 2y^2 = 0$$

Cela conduit aux solutions :

$$x = \pm\frac{1}{\sqrt{2}}$$

$$y = \pm\frac{1}{\sqrt{2}}$$

Optimisation sans contraintes

Exemple

Ainsi, les points stationnaires sont $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ et $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

Optimisation sans contraintes

Signe d'une matrice

- Attention : Ce n'est pas une condition suffisante : Un point critique peut être un minimum local, un maximum local, ou un bien un point de selle
- Un point critique $\mathbf{x}$ est un point de selle si pour tout $\varepsilon > 0$ il existe $\mathbf{a}, \mathbf{b} \in \mathcal{B}_{\varepsilon}(\mathbf{x})$ tels que $f(\mathbf{a}) < f(\mathbf{x}) < f(\mathbf{b})$
- Si $\mathbf{x}$ n'est pas un point critique, il ne peut pas être un minimum ou un maximum

Optimisation sans contraintes

Condition nécessaire de second ordre

- Si $\mathbf{x}^*$ est un minimum local de f sur $\mathbb{R}^n$, alors $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et la matrice hessienne $\nabla^2 f(\mathrm{x}^*)$ est semi-définie positive
- Preuve : Soit x^* un minimum local. Pour toute direction unitaire $\mathbf{d} \in \mathbb{R}^n$ et pour $t \in \mathbb{R}$ suffisamment petit :

$$
\begin{aligned}
f(\mathbf{x}^*) \leq f(\mathbf{x}^* + t\mathbf{d}) &\simeq f(\mathbf{x}^*) + t\mathbf{d}^\top \nabla f(\mathbf{x}^*) + \frac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\,\mathbf{d} \\
&= f(\mathbf{x}^*) + \frac{t^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\,\mathbf{d} \\
&\Rightarrow \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\,\mathbf{d} \geq 0
\end{aligned}
$$

- Si la matrice hessienne en un point critique est indéfinie, alors il s'agit d'un point de selle.

Optimisation sans contraintes

Condition suffisantes de second ordre

Théorème

Supposons que $\nabla^2 f$ *est continue dans un voisinage ouvert de* x^* *et que* $\nabla f(x^*) = 0$ *et* $\nabla^2 f(x^*)$ *est définie positive. Alors,* x^* *est un minimiseur local strict de* f.

Optimisation sans contraintes

Condition suffisantes de second ordre

Démonstration.

Supposons par contradiction que $\nabla f(x^*) > 0$. Définissons le vecteur $\mathbf{p} = -\nabla f(x^*)$ et notons que $\mathbf{p}^\top \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$. Puisque ∇f est continue près de x^*, il existe un scalaire $T > 0$ tel que $\mathbf{p}^\top \nabla f(x^* + t\mathbf{p}) < 0$ pour tout $t \in [0, T]$.
Pour tout $\bar{t} \in (0, T]$, nous avons, d'après le théorème de Taylor :

$$f(x^* + \bar{t}\mathbf{p}) \preceq f(x^*) + \bar{t}\mathbf{p}^\top \nabla f(x^* + tp), \text{ pour un certain } t \in (0, \bar{t}).$$

Par conséquent, $f(x^* + \bar{t}\mathbf{p}) < f(x^*)$ pour tout $\bar{t} \in (0, T]$. Nous avons trouvé une direction s'éloignant de x^* le long de laquelle f décroît, donc x^* n'est pas un minimiseur local, ce qui contredit notre hypothèse initiale. □

Optimisation sans contraintes

Condition suffisante d'optimalité du second ordre

Soit $f : U \to \mathbb{R}$ une fonction définie sur un ensemble ouvert $U \subset \mathbb{R}^n$. Supposons que f soit deux fois continûment différentiable sur U et que x^* soit un point stationnaire. Alors les résultats suivants sont vrais :

- Si la matrice hessienne évaluée en x^*, notée $H(x^*)$, est positive définie ($H(x^*) > 0$), alors x^* est un minimum local strict de f sur U.
- Si la matrice hessienne évaluée en x^*, notée $H(x^*)$, est négative définie ($H(x^*) < 0$), alors x^* est un maximum local strict de f sur U.

Optimisation sans contraintes

Point selle

Un point col ou point de selle pour une fonction $f : U \to \mathbb{R}$, définie sur un ensemble ouvert $U \subset \mathbb{R}^n$, est un point stationnaire où f est continûment différentiable et qui satisfait la condition suivante : x^* n'est ni un point de minimum local ni un point de maximum local pour f sur U. En d'autres termes, au point de selle, la fonction ne présente ni minimum local ni maximum local.

Optimisation sans contraintes

Condition suffisante pour un point de selle

Considérons une fonction $f : U \to \mathbb{R}$ définie sur un ensemble ouvert $U \subset \mathbb{R}^n$. Supposons que f soit deux fois continûment différentiable sur U et que x^* soit un point stationnaire. Si la matrice hessienne $H(x^*)$ est indéfinie ($\nabla^2 f(x^*)$ est une matrice indéfinie), alors x^* est un point de selle pour f sur U. En d'autres termes, si la matrice hessienne en x^* ne possède pas de caractéristiques claires de minimum ou de maximum, alors x^* est un point de selle pour f sur U.

Optimisation sans contraintes

Existence d'un minimum global

- Un autre problème important est de déterminer si une fonction possède effectivement un minimum global ou un maximum global.
- C'est la question de l'atteinte ou de l'existence. Un résultat très connu est dû à Weierstrass, affirmant qu'une fonction continue atteint son minimum et son maximum sur un ensemble compact.
- En d'autres termes, si la fonction est continue et que l'ensemble sur lequel elle est définie est compact, alors elle atteindra son minimum et son maximum sur cet ensemble.

Optimisation sans contraintes

Théorème de Weierstrass

Soit f une fonction continue définie sur un ensemble non vide et compact $C \subset \mathbb{R}^n$. Alors, il existe un point de minimum global pour f sur C et un point de maximum global pour f sur C. En d'autres termes, si la fonction est continue et définie sur un ensemble compact, elle possède à la fois un minimum global et un maximum global sur cet ensemble.

Optimisation sans contraintes

Théorème de Weierstrass

Soit f une fonction continue définie sur un ensemble non vide et compact $C \subset \mathbb{R}^n$. Alors, il existe un point de minimum global pour f sur C et un point de maximum global pour f sur C. En d'autres termes, si la fonction est continue et définie sur un ensemble compact, elle possède à la fois un minimum global et un maximum global sur cet ensemble.

Atteinte du Minimum et Maximum en l'Absence de Compacité.

Optimisation sans contraintes

Coercivité et Atteinte du Minimum Global

La fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite coercive si elle satisfait la propriété suivante :

$$\lim_{\|x\| \to \infty} f(x) = \infty$$

Une propriété essentielle des fonctions coercives, souvent utilisée dans ce contexte, est que toute fonction coercive atteint toujours un point de minimum global sur tout ensemble fermé.

Optimisation sans contraintes

Atteinte du Minimum sous Coercivité

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue et coercive, et soit $S \subset \mathbb{R}^n$ un ensemble non vide et fermé. Alors, f possède un point de minimum global sur S.

Optimisation sans contraintes

Exemple

Considérez la fonction $f(x_1, x_2) = x_1^2 + x_2^2$ sur l'ensemble $C = \{(x_1, x_2) : x_1 + x_2 < -1\}$.

Optimisation sans contraintes

Exemple

- L'ensemble C n'est pas borné, et donc le théorème de Weierstrass ne garantit pas l'existence d'un minimiseur global de f sur C.
- Cependant, comme f est coercitive et que C est fermé, le Théorème précedent garantit l'existence d'un tel minimiseur global.
- Il y a deux options : dans une option, le point de minimum global se trouve à l'intérieur de C. Dans ce cas, d'après le Théorème précedent, $\nabla f(x) = 0$, ce qui signifie que $x = 0$, ce qui est impossible car le vecteur nul n'est pas dans C.

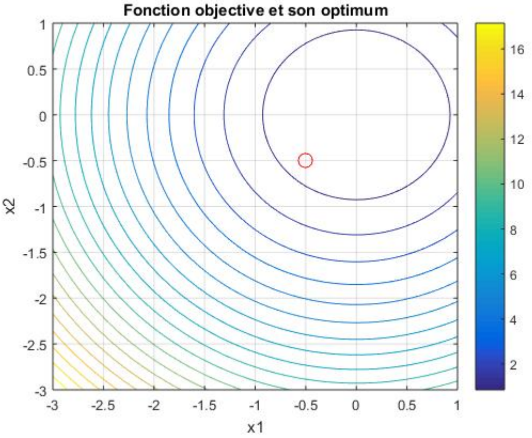
Optimisation sans contraintes

Exemple

- L'autre option est que le point de minimum global est atteint à la frontière de C donnée par $(C) = \{(x_1, x_2) : x_1 + x_2 = -1\}$. Nous pouvons alors substituer $x_1 = -x_2 - 1$ dans la fonction objectif et reformuler le problème comme un problème d'optimisation unidimensionnel consistant à minimiser $g(x_2) = (-1 - x_2)^2 + x_2$ sur $\mathbb{R}$.
- Puisque $g'(x_2) = 2(1 + x_2) + 2x_2$, il s'ensuit que g' a une seule racine, qui est $x_2 = -0.5$, et donc $x_1 = -0.5$.
- Comme $(x_1, x_2) = (-0.5, -0.5)$ est le seul candidat pour un point de minimum global, et comme il doit y avoir au moins un minimiseur global, il s'ensuit que $(x_1, x_2) = (-0.5, -0.5)$ est le point de minimum global de f sur C.

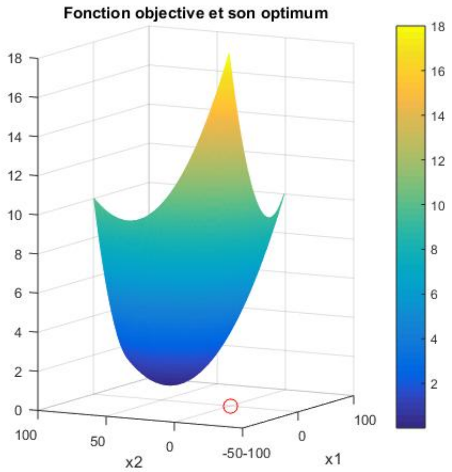
Optimisation sans contraintes

l'optimum



Optimisation sans contraintes

Illustration de la fonction



Optimisation sans contraintes

Minimum global

Theorem

Soit f une fonction deux fois continûment différentiable définie sur $\mathbb{R}^n$. Supposons que $\nabla^2 f(x) > 0$ pour tout $x \in \mathbb{R}^n$. Soit $x^* \in \mathbb{R}^n$ un point stationnaire de f. Alors, x^* est un point de minimum global de f.

Optimisation sans contraintes

Preuve

Grâce au théorème d'approximation linéaire, il en découle que pour tout $x \in \mathbb{R}^n$, il existe un vecteur $z_x \in [x^*, x]$ tel que

$$f(x) \geq f(x^*) + \nabla f(z_x)^\top (x - x^*).$$

Puisque $\nabla^2 f(z_x) > 0$, nous avons $f(x) > f(x^*)$, ce qui établit que x^* est un point de minimum global de f.

Optimisation sans contraintes

Exemple

Examinons en détail les calculs pour trouver les extremum de $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4$. ainsi que leur nature (maximum global, minimum global).Les dérivées partielles premières sont :

$$\frac{\partial f}{\partial x_1}(x, y) = -4x_1 + x_2^2 + 16x_1^3$$

$$\frac{\partial f}{\partial x_2}(x, y) = 2x_1x_2$$

Optimisation sans contraintes

Exemple

Nous devons résoudre le système d'équations $\frac{\partial f}{\partial x_1}(x, y) = 0$ et $\frac{\partial f}{\partial x_2}(x, y) = 0$. Commençons par $\frac{\partial f}{\partial x_1}(x, y) = 0$:

$$-4x_1 + x_2^2 + 16x_1^3 = 0$$

Ensuite, résolvons $\frac{\partial f}{\partial x_2}(x, y) = 0$:

$$2x_1x_2^2 = 0$$

Ainsi, les points critiques sont :
- $(0, 0)$.
- $\left(-\frac{1}{2}, 0\right)$.
- $\left(\frac{1}{2}, 0\right)$.

Optimisation sans contraintes

Exemple

Pour analyser la nature des points critiques à l'aide de la matrice hessienne $\nabla^2 f(x_1, x_2)$, nous devons calculer cette matrice et évaluer ses valeurs propres. Voici le calcul correct :
La matrice hessienne $\nabla^2 f(x_1, x_2)$ est donnée par :

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -4 + 48x_2^2 & 4x_1x_2 \\ 4x_1x_2 & 2x_2^2 \end{pmatrix}$$

Optimisation sans contraintes

Exemple

Évaluons cette matrice aux points critiques :

1. Pour $(0.5, 0)$, la matrice hessienne est $\nabla^2 f(0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix}$, ce qui est une matrice définie positive. Ainsi, $(0.5, 0)$ est un strict minimum local.

2. Pour $(-0.5, 0)$, la matrice hessienne est $\nabla^2 f(-0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & -1 \end{pmatrix}$, ce qui a des valeurs propres de signes opposés. Cela indique que $(-0.5, 0)$ est un point selle.

3. Pour $(0, 0)$, la matrice hessienne est $\nabla^2 f(0, 0) = \begin{pmatrix} -4 & 0 \\ 0 & 0 \end{pmatrix}$. Cette matrice n'est pas définie positive car elle a une valeur propre négative (-4) et une valeur propre nulle. Donc, $(0, 0)$ est également un point selle. C'est pourquoi $(0, 0)$ est un point selle..

Optimisation sans contraintes

Exemple 2

Le gradient de la fonction $f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$ est donné par :

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

Pour calculer les dérivées partielles, commençons par exprimer f en fonction de x_1 et x_2 :

$$f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$$

Optimisation sans contraintes

Exemple 2

Calculons maintenant les dérivées partielles :

$$\frac{\partial f}{\partial x_1} = 4x_1(x_1^2 + x_2^2 - 1)$$

$$\frac{\partial f}{\partial x_2} = 4x_2(x_1^2 + x_2^2 - 1) + 4x_2(x_2^2 - 1)$$

Les points stationnaires sont déterminés par les équations comme suit :

- $x_1 = 0$, et alors d'après l'équation (2.8), x_2 peut prendre l'une des valeurs 0, 1, ou -1.
- $x_1 + x_i = 1$, et alors d'après l'équation (2.8), x_2 peut être 0, ±1, et par conséquent x_1 peut être ±1, 0 respectivement.

Optimisation sans contraintes

Exemple 2

Ainsi, nous avons 5 points stationnaires : (0,0), (1,0), (-1,0), (0,1), et (0,-1). L'hessien de la fonction peut être calculé comme suit :

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

Nous pouvons calculer les dérivées partielles secondes, puis évaluer l'hessien en chaque point stationnaire pour déterminer leur nature (maximum, minimum, ou point col). Procédons à ce calcul.

Optimisation sans contraintes

Exemple 2

Maintenant que nous avons les points critiques $(0,0),(1,0),(-1,0),(0,1),(0,-1)$, nous pouvons analyser leur nature en utilisant la matrice hessienne. La matrice hessienne H pour la fonction $f(x_1,x_2)$ est donnée par :

$$H(x_1,x_2) = \begin{bmatrix} 3x_1^2 + x_2^2 - 1 & 2x_1x_2 \\ 2x_1x_2 & x_1^2 + 6x_2^2 - 2 \end{bmatrix}$$

Pour chaque point critique, nous évaluons la matrice hessienne et utilisons le critère de Sylvester pour déterminer la nature du point critique (minimum local, maximum local ou point selle).

Optimisation sans contraintes

Exemple 2

- Pour le point $(0,0)$ on a

$$H(0,0) = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$

- Pour le point $(1,0)$ on a

$$H(x_1,x_2) = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

- Pour le point $(-1,0)$ on a

$$H(x_1,x_2) = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

Optimisation sans contraintes

Exemple 2

- Pour le point $(0, 1)$ on a

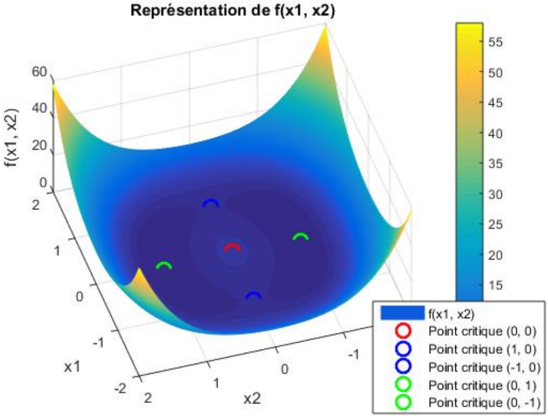
$$H(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

- Pour le point $(0, -1)$ on a

$$H(x_1, x_2) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

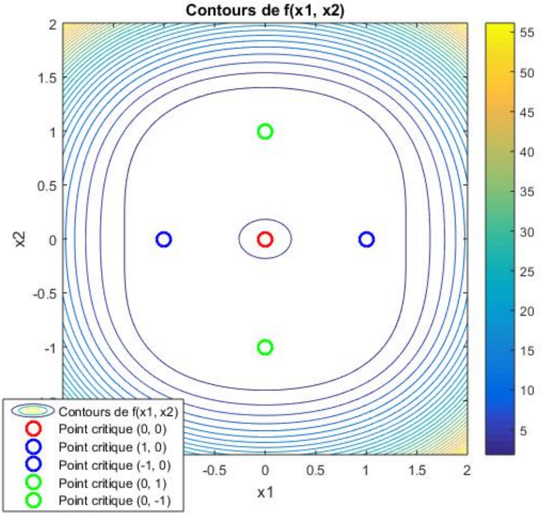
Optimisation sans contraintes

Exemple 2



Optimisation sans contraintes

Exemple 2



Fonctions quadratiques

Définition

Une fonction quadratique sur $\mathbb{R}^n$ est une fonction de la forme :

$$f(x) = x^T Ax + 2b^T x + c$$

où A est une matrice symétrique dans $\mathbb{R}^{n\times n}$, b est un vecteur dans $\mathbb{R}^n$, et c est un scalaire dans $\mathbb{R}$.

Fonctions quadratiques

Définition

Nous ferons souvent référence à la matrice A dans l'équation précédente comme la matrice associée à la fonction quadratique f. Le gradient et le hessien d'une fonction quadratique ont des formules analytiques simples :

$$\nabla f(x) = 2Ax + 2b,$$
$$\nabla^2 f(x) = 2A.$$

Par les formules ci-dessus, nous pouvons déduire plusieurs propriétés importantes des fonctions quadratiques qui sont associées à leurs points stationnaires.

Fonctions quadratiques

Théorème

Soit $f(x) = x^T Ax + 2b^T x + c$, où $A \in \mathbb{R}^{n\times n}$ est symétrique, $b \in \mathbb{R}^n$, et $c \in \mathbb{R}$. Alors :

- x est un point stationnaire de f si et seulement si $Ax = -b$,
- si $A \succeq 0$ (c'est-à-dire, A est définie positive), alors x est un point de minimum global de f si et seulement si $Ax = -b$,
- si $A \succ 0$, alors $x = -A^{-1}b$ est un point de minimum global strict de f.

Fonctions quadratiques

Coercivité des fonctions quadratiques

Lemma

Soit $f(x) = x^T Ax + 2b^T x + c$*, où* $A \in \mathbb{R}^{n\times n}$ *est symétrique,* $b \in \mathbb{R}^n$*, et* $c \in \mathbb{R}$*. Alors,* f *est coercive si et seulement si* $A \succ 0$*.*

Contents

Optimisation sans contraintes

Convexité

Démonstration.

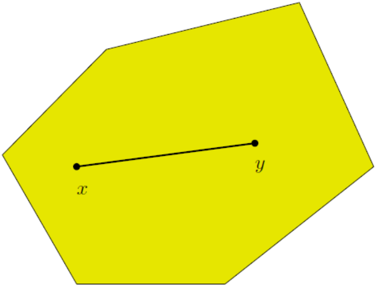
Lorsque la fonction objectif est **convexe**, les minimiseurs locaux et globaux sont simples à caractériser. □

Optimisation sans contraintes

Convexité

Un ensemble C est *convexe* si, pour tous les points $\mathbf{x}, \mathbf{y}$ dans C et tous λ dans $[0, 1]$, le point $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ est aussi dans C. Cela peut être exprimé par la formule :

$$\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in C, \quad \forall \mathbf{x}, \mathbf{y} \in C, \forall \lambda \in [0, 1]$$



Optimisation sans contraintes

Convexité

Voici quelques exemples d'ensembles convexes :

- L'ensemble des points dans le plan est convexe.
- L'ensemble vide $\emptyset$ est considéré convexe par convention.
- Une droite est un ensemble convexe.
- Un demi-espace est un ensemble convexe. Par exemple, l'ensemble $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} \leq b\}$ est un demi-espace défini par le vecteur normal $\mathbf{a}$ et la constante b.
- Un polyèdre est un ensemble convexe. Par exemple, un tétraèdre dans $\mathbb{R}^3$ est un polyèdre convexe.
- L'ensemble $\mathbb{R}^n$ lui-même est convexe.

Optimisation sans contraintes

Convexité

Théorème :

- L'intervalle $I = [a, b]$ est un ensemble convexe dans $\mathbb{R}$.
- L'intersection d'ensembles convexes est convexe.
- L'union d'ensembles convexes est convexe.

Optimisation sans contraintes

Fonction convexe

Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est convexe si, pour tous x, y dans son domaine et pour tout t dans l'intervalle $[0, 1]$, elle satisfait l'inégalité :

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

En termes plus simples, une fonction est convexe si le segment de droite entre deux points choisis sur sa courbe ou au-dessus de celle-ci reste toujours au-dessus de la courbe.

Optimisation sans contraintes

Exemples de fonctions convexes

1. **Fonctions linéaires** : Toute fonction linéaire $f(x) = ax + b$, où a et b sont des constantes, est convexe.
2. **Fonctions quadratiques** : Les fonctions de la forme $f(x) = x^2$ ou $f(x) = x^2 + ax + b$ sont convexes sur leur domaine.
3. **Fonctions exponentielles** : Les fonctions de la forme $f(x) = e^{ax}$ sont convexes pour toute constante réelle a.
4. **Normes** : Les fonctions comme $f(x) = \|x\|$, où $\|x\|$ est une norme (par exemple, la norme euclidienne $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$), sont convexes.

Optimisation sans contraintes

Exemples de fonctions convexes

1. **Fonction indicatrice convexe :** La fonction $f(x)$ définie par :

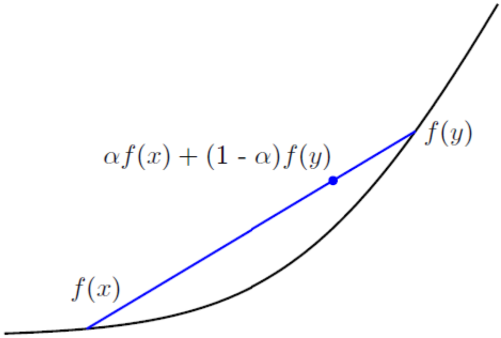
$$f(x) = \begin{cases} 0 & \text{si } x \text{ est dans un ensemble convexe } C \\ +\infty & \text{sinon} \end{cases}$$

   est convexe, où C est un ensemble convexe.

2. **Maximum de fonctions convexes :** Si f_i est convexe pour $i = 1, 2, ..., n$, alors $f(x) = \max\{f_1(x), f_2(x), ..., f_n(x)\}$ est convexe.
   Ce ne sont que quelques exemples courants, mais la classe des fonctions convexes est assez vaste et inclut de nombreuses autres fonctions et combinaisons de fonctions qui satisfont la définition de convexité.

Optimisation sans contraintes

Représentation graphique



Optimisation sans contraintes

Le sous-gradient

Un vecteur g dans $\mathbb{R}^n$ est un sous-gradient de f en $x \in X$ si, pour tout $y \in X$, la condition suivante est satisfaite :

$$f(y) \geq f(x) + \langle g, y - x \rangle.$$

L'ensemble des sous-gradients de f en x est noté $\partial f(x)$.

Optimisation sans contraintes

Le sous-gradient

Bien sûr, voici quelques exemples de sous-gradients dans différents contextes mathématiques courants :

1. **Fonction Convexe** : - Pour une fonction convexe f dans $\mathbb{R}$, le sous-gradient en un point x est simplement sa dérivée (si elle est dérivable à ce point).

2. **Norme L1** : - Pour la norme ℓ_1 d'un vecteur x dans $\mathbb{R}^n$, le sous-gradient est donné par $\partial||x||_1 = \text{sign}(x)$, où $\text{sign}(x)$ est le vecteur dont chaque composante est le signe de la composante correspondante de x.

3. **Fonction Maximum** : - Pour la fonction maximum $f(x) = \max\{f_1(x), f_2(x)\}$ de deux fonctions f_1 et f_2, le sous-gradient en un point x est le sous-gradient de f_1 en x si $f_1(x) > f_2(x)$, et le sous-gradient de f_2 en x si $f_2(x) > f_1(x)$.

Optimisation sans contraintes

Le sous-gradient

Propositions :

- Si $\partial f(x) \neq \emptyset$ pour tous les $x \in X$, alors f est convexe.
- Si f est convexe, alors pour chaque $x \in \mathcal{X}$, $\partial f(x) \neq \emptyset$.
- Si f est convexe et différentiable en x, alors $\partial f(x) = \{\nabla f(x)\}$, où $\nabla f(x)$ est le gradient de f en x.

Ces énoncés établissent des liens importants entre la convexité d'une fonction, l'existence de sous-gradients et le cas particulier où la fonction est différentiable.

Optimisation sans contraintes

Le sous-gradient

- Supposons que f n'est pas convexe. Cela signifie qu'il existe $x, y \in X$ et $t \in [0, 1]$ tels que $f(tx + (1-t)y) > tf(x) + (1-t)f(y)$. Maintenant, considérons $\partial f(tx + (1-t)y)$. Comme f n'est pas convexe, $\partial f(tx + (1-t)y) = \emptyset$, ce qui est une contradiction.
- Supposons par l'absurde que f est convexe mais il existe un x tel que $\partial f(x) = \emptyset$. Cela signifie que f n'est pas différentiable en x. Cependant, une fonction convexe est différentiable presque partout. Donc, c'est une contradiction.
- Si f est différentiable en x, alors le sous-gradient est réduit au gradient. Formellement, $\partial f(x) = \{\nabla f(x)\}$. Cela découle de la définition du sous-gradient et du fait que pour une fonction différentiable, le gradient est le seul sous-gradient possible. La convexité assure que le gradient est un sous-gradient.

Optimisation sans contraintes

L'inégalité du gradient

Theorem

Soit $f : C \to \mathbb{R}$ *une fonction continûment différentiable définie sur un ensemble convexe* $C \subset \mathbb{R}^n$. *Alors,* f *est convexe sur* C *si et seulement si*

$$f(x) + \nabla f(x)^\top (y - x) \leq f(y)$$

pour tout $x, y \in C$.

Optimisation sans contraintes

Fonction strictement convexes

Proposition

Soit $f : \mathbb{C} \to \mathbb{R}$ une fonction continûment différentiable définie sur un ensemble convexe $C \subseteq \mathbb{R}^n$. Alors f est strictement convexe sur C si et seulement si

$$f(x) + \nabla f(x)^T (y - x) < f(y)$$

pour tout $x, y \in C$ tels que $x \neq y$.

Optimisation sans contraintes

Convexité des fonctions quadratiques

Théorème

Soit $f : \mathbb{R}^n \to \mathbb{R}$ *la fonction quadratique donnée par* $f(x) = x^T Ax + 2b^T x + c$ *où* $A \in \mathbb{R}^{n\times n}$ *est symétrique,* $b \in \mathbb{R}^n$ *et* $c \in \mathbb{R}$*.*
Alors f *est (strictement) convexe si et seulement si* $A \succeq 0$ *(*$A \succ 0$*).*

Optimisation sans contraintes

Convexité des fonctions quadratiques

Démonstration.

La convexité de f est équivalente à

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \text{ pour tout } x, y \in \mathbb{R}^n.$$

Cela est similaire à

$$y^T Ay + 2b^T y + c \geq x^T Ax + 2b^T x + c + 2(Ax + b)^T (y - x) \text{ pour tout } x, y \in \mathbb{R}^n.$$

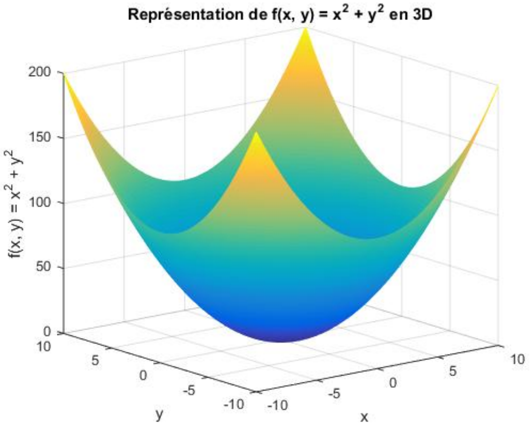
$$(y - x)^T A(y - x) \geq 0 \text{ pour tout } x, y \in \mathbb{R}^n.$$

Cela est équivalent à l'inégalité $d^T Ad \geq 0$ pour tout $d \in \mathbb{R}^n$. C'est la même chose que $A \succeq 0$. Des arguments similaires montrent que la convexité stricte est équivalente à $d^T Ad > 0$ pour tout $d \in \mathbb{R}^n$ non nul, c'est-à-dire $A \succ 0$. $\square$

Optimisation sans contraintes

Exemples des fonctions quadratiques

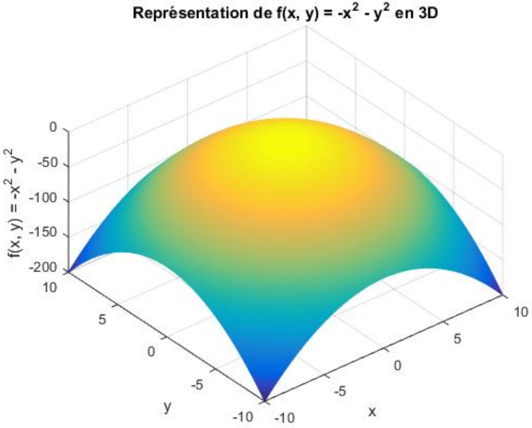
Fonction convexe : $f(x, y) = x^2 + y^2$



Optimisation sans contraintes

Exemples des fonctions quadratiques

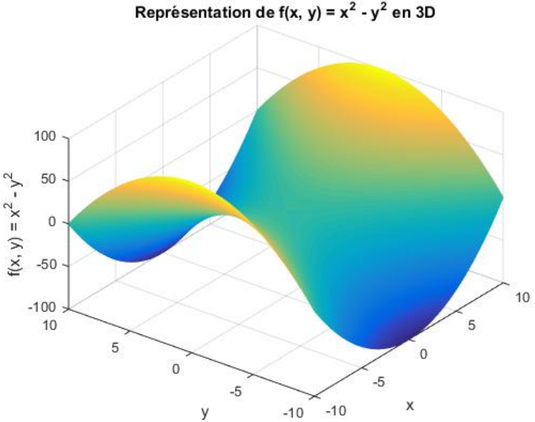
Fonction concave : $f(x, y) = -x^2 - y^2$



Optimisation sans contraintes

Exemples des fonctions quadratiques

Fonction ni convexe ni concave : $f(x, y) = x^2 - y^2$



Optimisation sans contraintes

Monotonie du gradient

Théorème

Suppose that f is a continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$$

for any $x, y \in C$.