Chapitre 3 : Optimisation non-linéaire sans contraintes

3.1 Introduction

De nombreuses applications industrielles de l'optimisation impliquent une modélisation plus proche de la physique. En optimisation de la conception par exemple, on optimise des méta-modèles ou surfaces de réponse issus d'un modèle statistique qui ne sont pas linéaires. De même l'optimisation topologique ou l'optimisation de forme fait largement appel à des modèles non linéaires et de plus de très grande taille dans cet exemple. Formellement un programme mathématique non linéaire sans contraintes s'écrit :

$$\begin{array}{l} Min\ f(x) \\ \mathrm{x} \in \mathrm{S} \end{array}$$

x vecteur de nombre réels représentant les variables de décision

Exemple 3.1 : Déterminer les valeurs extrémales de $\begin{array}{l} Min\ f(x) = x^2 \\ \mathrm{x} \in \mathrm{R} \end{array}$

Les techniques pour résoudre les problèmes mathématiques et la performance des algorithmes d'optimisation dépendent du nombre de variable de décision, et la nature de la fonction objectif.

Pour les fonctions dérivables deux fois, des problèmes sans contraintes peuvent être résolus en trouvant les points où le gradient de la fonction est 0 et en utilisant la matrice Hessienne pour classer la nature de point. Si le Hessien est défini positif, le point est un minimum local ; s'il est un défini négatif, un maximum local.

Si la fonction est convexe sur l'ensemble des solutions admissibles alors tout minimum local est aussi un minimum global.

3.2 Rappel mathématiques

Rappel mathématiques 1: le gradient d'une fonction f à n variables est un vecteur n-dimensionnel contenant toutes les dérivées partielles de f par rapport à chacune de ses variables

$\nabla f(x1, x2,..., xn) = (\frac{\partial f}{x1}, \frac{\partial f}{x2},..., \frac{\partial f}{xn})$

Si $\nabla f(x1, x2,..., xn) = 0$ pour le point (x_1', x_2',.., x_n') alors ce point est un minimum (maximum) local. Ce point est appelé également point stationnaire.

Rappel mathématiques 2: Le **Hessien** est une façon d'écrire les dérivées d'un gradient à l'aide d'une matrice symétrique. pour une fonction f(x1,x2,…,xn), le Hessein est calculé en utilisant la formule suivante :

$H(x1,, x2,..., xn) = \nabla^2 f(x1,, x2,..., xn) = \left(\frac{\partial f}{\partial xi \partial xj}\right)$ pour i=1..n,j=1..n.

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

Pour le cas d'une fonction à deux variables : Soit $f(x, y)$ une fonction deux fois dérivable, alors le Hessien est :

$$H(x, y) = \nabla^2 f(x, y) = \begin{bmatrix} f_{xx}^{(2)}(x, y) & f_{xy}^{(2)}(x, y) \\ f_{yx}^{(2)}(x, y) & f_{yy}^{(2)}(x, y) \end{bmatrix}.$$

On a toujours $f_{xy}^{(2)}(x, y) = f_{yx}^{(2)}(x, y)$.

Exemple 3.2 $f(x, y) = 5x^2 + 3y^2 + 8xy - 3x + 7y$ où $(x, y) \in \Re^2$.

Calculer le gradient de f : $f_x'(x, y) = 10x + 8y - 3$; $f_y'(x, y) = 6y + 8x + 7$.

Calculer le Hessien de f : $\frac{\partial}{\partial x} f_x'(x, y) = f_{xx}^{(2)}(x, y) = 10$; $\frac{\partial}{\partial y} f_x'(x, y) = f_{xy}^{(2)}(x, y) = 8$;

$\frac{\partial}{\partial x} f_y'(x, y) = f_{yx}^{(2)}(x, y) = 8$; $\frac{\partial}{\partial y} f_y'(x, y) = f_{yy}^{(2)}(x, y) = 6$.

Le Hessien est $H(x,y)=\begin{bmatrix}10 & 8\\ 8 & 6\end{bmatrix}$ et dans cet exemple, il est le même (constant) en tout point du domaine de $f(x,y)$, ce qui n'est pas toujours le cas.

- Soit $A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$, le **déterminant** de A est donné par $|A|=\det(A)=ad-bc$.
- Une matrice A est **symétrique** si $b = c$; le Hessien est toujours symétrique.

3.3 Définitions

Soit A une matrice n*n symétrique et X un vecteur de R^n X^T est le transposé de X.

- A est **définie positive** si et seulement si $X^T.A.X > 0 \ \forall \ X \in S$

  Pour le cas d'une fonction f à deux variables il faut vérifier seulement $\det(A) > 0, a > 0, d > 0$.
- A est **définie négative** si et seulement si $X^T.A.X < 0$

  Pour le cas d'une fonction f à deux variables il faut vérifier seulement $\det(A) > 0, a < 0, d < 0$.
- A est **semi-définie positive** si et seulement si $X^T.A.X \geq 0$

  Pour le cas d'une fonction f à deux variables il faut vérifier seulement $\det(A) \geq 0, a \geq 0, d \geq 0$
- A est **semi-définie négative** si et seulement si $X^T.A.X \leq 0$

  Pour le cas d'une fonction f à deux variables il faut vérifier seulement $\det(A) \geq 0, a \leq 0, d \leq 0$
- Note : $\det(H(x,y)) \geq 0$ signifie que le déterminant $\det(H(x,y))$ peut s'annuler pour une ou plusieurs valeurs de (*x*,*y*).

Exemples 3.3

- $H_1=\begin{bmatrix}10 & 8\\ 8 & 6\end{bmatrix} \rightarrow \det(H_1)=10\cdot 6-8\cdot 8=-4<0$; H_1 est non définie.
- $H_2=\begin{bmatrix}-2 & 3\\ 3 & -6\end{bmatrix} \rightarrow \begin{cases}\det(H_2)=(-2)(-6)-3\cdot 3=3>0\\ \text{diagonale négative}: a=-2;\ d=-6.\end{cases}$ H_2 est définie négative.
- $H_3=\begin{bmatrix}10 & 10\\ 10 & 10\end{bmatrix} \rightarrow \begin{cases}\det(H_3)=10\cdot 10-10\cdot 10=0\\ \text{diagonale positive}: a=10,\ d=10\end{cases}$ H_3 est semi–définie positive.
- $H_4(x,y)=\begin{bmatrix}x^2 & 0\\ 0 & y^2\end{bmatrix} \rightarrow \begin{cases}\det(H_4(x,y))=x^2y^2 \geq 0\\ a=x^2 \geq 0; b=x^2 \geq 0\end{cases}$ H_4 est semi–définie positive.

- $H_5(x,y) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \begin{cases} \det(H_5(x,y)) = 0 \\ a = b = 0 \end{cases}$

- H_5 est à la fois semi–définie positive et semi-définie négative; c'est en fait le Hessien d'une fonction linéaire.

Théorème : Soit $X^* = (x^*, y^*)$ un point stationnaire de la fonction $f(x,y)$.

- **Si le Hessien de *f* est défini positif en X^*, alors X^* est un *minimum local.***
- **Si le Hessien de *f* est défini négatif en X^*, alors X^* est un *maximum local.***
- **Si $\det(H(X^*)) < 0$, alors X^* est un *point singulier* .**
- **Si $\det(H(X^*)) = 0$, il faut utiliser d'autres techniques basées sur les dérivées d'ordre supérieur pour pouvoir se prononcer.**

Exemple 3.4 $f(x,\ y) = 2x^2 - 8xy^2 - \dfrac{16}{3}y^3 + 8y^2 - 8,$ où $(x,y) \in \Re^2$.

a) Donner le **gradient** $\nabla f(x,y)$ et le **Hessien** $H(x,y)$.

b) Identifier la nature des points stationnaires de la fonction $f(x,y)$.

a) $\nabla f(x,y) = (4x - 8y^2; -16xy - 16y^2 + 16y)$ **;** $H(x,y) = \begin{bmatrix} 4 & -16y \\ -16y & -16x - 32y + 16 \end{bmatrix}$**.**

b) $\nabla f(x,y) = (0;0) \Leftrightarrow x = 2y^2$ et $xy + y^2 - y = 0 = 2y^3 + y^2 - y = y(2y-1)(y+1)$ **.**

Les points stationnaires sont : (0, 0) (1/2, 1/2) (2, -1).

$$H(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}; \quad \det(H(0,0)) > 0; \quad 4 > 0 \qquad \Rightarrow \qquad (0,0) \text{ est un minimum local.}$$

$$H(\frac{1}{2},\frac{1}{2}) = \begin{bmatrix} 4 & -8 \\ -8 & -8 \end{bmatrix}; \quad \det(H(\frac{1}{2},\frac{1}{2})) < 0 \qquad \Rightarrow \qquad (\frac{1}{2},\frac{1}{2}) \text{ est un point singulier.}$$

$$H(2,-1) = \begin{bmatrix} 4 & 16 \\ 16 & 16 \end{bmatrix}; \quad \det(H(2,-1)) < 0; \qquad \Rightarrow \qquad (2,-1) \text{ est un point singulier.}$$

Exemple 3.5

Soit la fonction de profit suivante $f(x,y) = 300x + 400y - x^2 - 2y^2 - 2xy$

Cette fonction possède un seul point stationnaire $(x^*, y^*) = (100,\ 50)$.

Quelle est la nature de ce point ?

Puisque $\nabla f(x,y) = (300 - 2x - 2y; 400 - 4y - 2x)$, le Hessien est $H(x,y) = \begin{bmatrix} -2 & -2 \\ -2 & -4 \end{bmatrix}$.

$\det(H(100, 50)) = (-2)(-4) - (-2)(-2) = 4 > 0$ avec $a = -2$ et $b = -4$. $H(x^*, y^*)$ est *défini négatif*. $(x^*, y^*) = (100, 50)$ est donc un *maximum local*.

Pour une fonction convexe ou concave, la condition du premier ordre (gradient nul) est nécessaire et suffisante pour caractériser la nature d'un point stationnaire.

Théorème : Soit $X^* = (x^*, y^*)$ **un point stationnaire de la fonction** $f(x,y)$**.**

- **Si** $f(x,y)$ **est convexe, alors** X^* **est un minimum absolu (global).**
- **Si** $f(x,y)$ **est concave, alors** X^* **est un maximum absolu (global).**

Si la fonction $f(x,y)$ est deux fois dérivable, on peut la caractériser par le Hessien :

Théorème : Soit $f(x,y)$ **dérivable deux fois sur un ensemble** E**.**

$f(x,y)$ **est convexe si et seulement si** $H(x,y)$ **est semi-défini positif** $\forall (x,y) \in E$**.**

$f(x,y)$ **est concave si et seulement si** $H(x,y)$ **est semi-défini négatif** $\forall (x,y) \in E$**.**

Si $\det(H(x,y)) > 0, \ \forall (x,y) \in E$ **(H est défini), alors on dira que la fonction** $f(x,y)$ **est** *strictement* **convexe ou** *strictement* **concave.**

Exemple 3.6 $f(x,y) = x\ln(x) + y\ln(y), \quad x > 0, y > 0$.

$\nabla f(x,y) = (1 + \ln(x), \quad 1 + \ln(y)), \quad x > 0, y > 0$.

$H(x,y) = \begin{bmatrix} 1/x & 0 \\ 0 & 1/y \end{bmatrix}, \quad x > 0, y > 0. \qquad \det(H(x,y)) = (1/xy) > 0; \;$ et $\; 1/x > 0, \forall (x,y).$

$H(x,y)$ est défini positif $\forall (x,y)$, donc $f(x,y)$ est *strictement* **convexe.**

Exemple 3.7 $f(x,y) = 4x^2 + 2y^2 - 8x - 2y + 1$.

Identifier la nature des points stationnaires de la fonction $f(x,y)$.

$$\nabla f(x,y) = (8x-8; \quad 4y-2)\,; \quad \nabla f(x^*,y^*) = 0 \Leftrightarrow \begin{cases} 8x^* - 8 = 0 \Leftrightarrow 8x^* = 8 \Rightarrow x^* = 1 \\ 4y^* - 2 = 0 \Leftrightarrow 4y^* = 2 \Rightarrow y^* = 1/2 \end{cases}$$

(1 , 1/2) est le seul point stationnaire de $f(x,y)$.

$H(x,y) = \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix}$. $\det(H(x^*,y^*)) = 8 \cdot 4 - 0 \cdot 0 = 32 > 0$ avec $a = 8$, $\forall(x, y)$.

Le Hessien $H(X^*)$ est défini positif $\forall(x, y)$, donc $f(x,y)$ est *strictement* **convexe.**

Par conséquent, $(x^*, y^*) = (1, 1/2)$ est un *minimum absolu*.

Suite de l'exemple 3.5

On a déjà vu que l'unique point stationnaire $(x^*, y^*) = (100, 50)$ est un maximum local. Mais il est possible d'avoir une information plus complète sur ce point stationnaire, puisque le Hessien $H(x,y) = \begin{bmatrix} -2 & -2 \\ -2 & -4 \end{bmatrix}$ est le même partout.

$\det(H) = (-2)(-4) - (-2)(-2) = 4 > 0$ avec $a = -2$ et $b = -4$. $H(x,y)$ est *défini négatif* pour tout point (x,y) et donc $f(x,y)$ est *strictement* **concave**. Par conséquent, $(x^*, y^*) = (100, 50)$ est un *maximum absolu*.

Propriétés des fonctions convexes et concaves

- Si f_1 et f_2 sont convexes (concaves), alors $f_1 + f_2$ est aussi convexe (concave).
- Si *f* est convexe (concave), alors $g(X) = k \cdot f(X)$ est convexe (concave), $\forall k > 0$.
- Si *f* est convexe (concave) alors $-f$ est concave (convexe).
- $f(x,y) = a + bx + cy$ est dite une **fonction affine** sur $\Re^2$; elle est à la fois convexe et concave parce que son Hessien est nul et qu'il est alors à la fois semi-défini positif et semi-défini négatif.

Références

1. Initiation aux techniques classiques de l'optimisation, Michèle Breton et Alain Haurie, Éditeur Modulo.
2. Calcul 1 : Théorie, exemples, problèmes, Gilles Ouellet, Éditions Le Griffon d'Argile (Révision sur les dérivées).
3. Initiation aux techniques classiques de l'optimisation - solutions, Michèle Breton et Alain Haurie, Éditeur Modulo.