Chapitre 4

OPTIMISATION SANS CONTRAINTES.CONDITIONS D'OPTIMALITE

4.1 Définitions

Définition(Fonction convexe différentiable) Soit $C \sqsubseteq \mathbb{R}^n$ non vide, $f : \mathbb{R}^n \to \mathbb{R}$ et $\hat{x} \in int(C).f$ est dite différentiable au point $\hat{x}$, s'il existe un vecteur $A \in \mathbb{R}^n$et une

fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ telle que

$$f\left(x\right) = f\left(\hat{x}\right) + A(x - \hat{x}) + \left\|x - \hat{x}\right\| \alpha(\hat{x}, x - \hat{x}),$$

où $\alpha(\hat{x}, x - \hat{x})_{x \to \hat{x}} \to 0$.On peut note le vecteur A comme suit : $A = \nabla f\left(\hat{x}\right)^T$.

Définition(Fonction convexe deux foix différentiable) Soit $C \sqsubseteq \mathbb{R}^n$ non vide, $f : \mathbb{R}^n \to \mathbb{R}$ est dite deux foix différentiable ou point $\hat{x} \in int(C)$ s'il existe un vecteur $\nabla f\left(\hat{x}\right)$ et une matrice symétrique $H(\hat{x})$ d'ordre $(n \times n)$ appellèe matrice hessienne, et une fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ telle que

$$\forall x \in C, f(x) = f\left(\hat{x}\right) + \nabla f\left(\hat{x}\right)^T (x - \hat{x}) + \frac{1}{2}(x - \hat{x})^T H(\hat{x})(x - \hat{x}) + \left\| x - \hat{x} \right\|^2 \alpha(\hat{x}, x - \hat{x}),$$

où $\alpha(\hat{x}, x - \hat{x})_{x \to \hat{x}} \longrightarrow 0$ et $H(\hat{x})$ est la matrice Hessienne au point $\hat{x}$.

Définition Soit $f : \mathbb{R}^n \to \mathbb{R}$, on appelle problème de minimisation sans contraites le problème (P) suivant :

$$(P) \quad \min \{f(x) : x \in \mathbb{R}^n\}.$$

1) $\hat{x} \in \mathbb{R}^n$ est un minimum global de (P) si et seulement si

$$f\left(\hat{x}\right) \leq f(x) : \forall x \in \mathbb{R}^n.$$

2) $\hat{x} \in \mathbb{R}^n$ est un minimum local de (P) si et seulement si il existe un voisinage $V_\varepsilon\left(\hat{x}\right)$ tel que

$$f\left(\hat{x}\right) \leq f(x) : \forall x \in V_\varepsilon\left(\hat{x}\right).$$

3) $\hat{x} \in \mathbb{R}^n$ est un minimum local strict de (P) si et seulement si il existe un voisinage $V_\varepsilon\left(\hat{x}\right)$ tel que

$$f\left(\hat{x}\right) < f(x) : \forall x \in V_\varepsilon\left(\hat{x}\right), \; x \neq \hat{x}.$$

4.2 Direction de descente

Définition Soit $f : \mathbb{R}^n \to \mathbb{R}$, $\hat{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ est dite direction de descente au point $\hat{x}$ si et seulement si il existe $\delta > 0$ tel que

$$f\left(\hat{x} + \lambda d\right) < f\left(\hat{x}\right) : \forall \lambda \in \left]0, \delta\right[.$$

Donnons maintenant une condition suffisante pour que d soit une direction de descente.

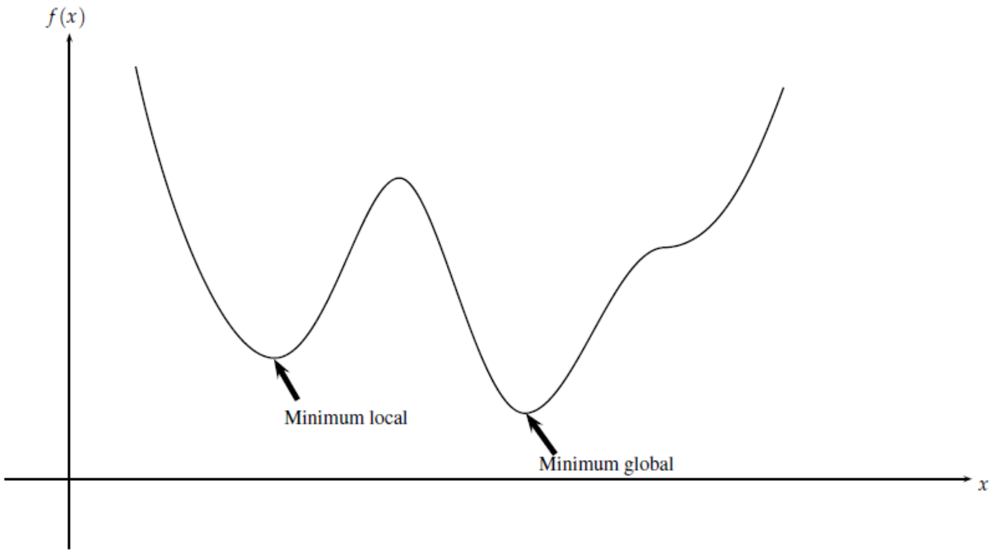

Figure 2.1: Les différents types de minimas.

Théorème Soit $f : \mathbb{R}^n \to \mathbb{R}$ differentiable au point $\hat{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ une direction vérifiant la condition suivante :

$$f' \left(\hat{x}, d\right) = \nabla f \left(\hat{x}\right)^t \cdot d < 0.$$

Alors d est une direction de descente au point $\hat{x}$.

Preuve. f est différentiable au point $\hat{x}$. Donc

$$f \left(\hat{x} + \lambda d\right) = f \left(\hat{x}\right) + \lambda \nabla f \left(\hat{x}\right)^t \cdot d + \lambda \|d\| \alpha \left(\hat{x}, \lambda d\right),$$

avec

$$\alpha \left(\hat{x}, \lambda d\right) \underset{\lambda \to 0}{\to} 0,$$

ceci implique que

$$f' \left(\hat{x}, d\right) = \lim_{\lambda \to 0} \frac{f \left(\hat{x} + \lambda d\right) - f \left(\hat{x}\right)}{\lambda} = \nabla f \left(\hat{x}\right)^t \cdot d < 0.$$

La limite étant strictement négative, alors il existe un voisinage de zéro $V(0) =]-\delta, +\delta[$ tel que

$$\frac{f \left(\hat{x} + \lambda d\right) - f \left(\hat{x}\right)}{\lambda} < 0, \quad \forall \lambda \in]-\delta, +\delta[\,. \tag{1}$$

La relation (1) est particulièrement vraie pour tout $\lambda \in]0, +\delta[$. On obtient le résultat cherché en multipliant la relation (1) par $\lambda > 0$. ∎

4.3 Schémas général des algorithmes d'optimisation sans contraintes

Supposons que d_k soit une direction de descente au point x_k. Ceci nous permet de considérer le point x_{k+1}, successeur de x_k, de la manière suivante :

$$x_{k+1} = x_k + \lambda_k d_k, \quad \lambda_k \in \left]0, +\delta\right[.$$

Vu la définition de direction de descente, on est assuré que

$$f(x_{k+1}) = f(x_k + \lambda_k d_k) < f(x_k).$$

Un bon choix de d_k et de λ_k permet ainsi de construire une multitude d'algorithmes d'optimisation.

4.3.1 Exemples de choix de directions de descente

Par exemple si on choisit $d_k = -\nabla f(x_k)$ et si $\nabla f(x_k) \neq 0$, on obtient la méthode du gradient. La méthode de Newton correspond à $d_k = -(H(x_k))^{-1} \cdot \nabla f(x_k)$. Bien sur $-\nabla f(x_k)$ est une direction de descente $(\nabla f(x_k)^t \cdot d_k = -\nabla f(x_k)^t \cdot \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0)$. Pour la deuxième direction si la matrice hessienne $H(x_k)$ est définie positive alors $d_k = -(H(x_k))^{-1} \cdot \nabla f(x_k)$ est aussi une direction de descente.

4.3.2 Exemple de choix de pas λ_k

On choisit en général λ_k de façon optimale, c'est à dire que λ_k doit vérifier

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k) : \ \forall \lambda \in \left[0, +\infty\right[.$$

En d'autres termes on est ramené à étudier à chaque itération un problème de minimisation d'une variable réelle. C'est ce qu'on appelle recherche linéaire.

4.4 Conditions nécessaires d'optimalité

4.4.1 Conditions nécessaires d'optimalité du premier ordre

Théorème Soit $f : \mathbb{R}^n \to \mathbb{R}$ differentiable au point $\hat{x} \in \mathbb{R}^n$.Si $\hat{x}$ est un minimum local de (P) alors $\nabla f\left(\hat{x}\right) = 0$.

Preuve. C'est une conséquence directe duThéorème 1. En effet, supposons que $\nabla f\left(\hat{x}\right) \neq 0$. Puisque la direction $d = -\nabla f\left(\hat{x}\right)$ est une direction de descente, alors il existe $\delta > 0$ tel que :

$$f\left(\hat{x} + \lambda d\right) < f\left(\hat{x}\right), \forall \in \left]0, \delta\right[.$$

Ceci est contradiction avec le fait que $\hat{x}$ est une solution optimale locale de (P) .■ ■

4.4.2 Condition nécessaire d'optimalité du second ordre

Théorème Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois differentiable au point $\hat{x} \in \mathbb{R}^n$.Si $\hat{x}$ est un minimum local de (P) alors $\nabla f\left(\hat{x}\right) = 0$ et la matrice hessienne de f au point $\hat{x}$, qu'on note $H\left(\hat{x}\right)$, est semi définie positive.

Preuve. Soit $x \in \mathbb{R}^n$ quelconque, f étant deux fois différentiable au point $\hat{x}$, on aura pour tout $\lambda \neq 0$

$$f\left(\hat{x} + \lambda x\right) = f\left(\hat{x}\right) + \frac{1}{2}\lambda^2 x^t H\left(\hat{x}\right) x + \lambda^2 \left\|x^2\right\| \alpha\left(\hat{x}, \lambda x\right), \qquad \alpha\left(\hat{x}, \lambda x\right) \underset{\lambda \to 0}{\to} 0.$$

Ceci implique

$$\frac{f\left(\hat{x} + \lambda x\right) - f\left(\hat{x}\right)}{\lambda^2} = \frac{1}{2} x^t H\left(\hat{x}\right) x + \left\|x^2\right\| \alpha\left(\hat{x}, \lambda x\right). \tag{2}$$

$\hat{x}$ est un optimum local, il existe alors $\delta > 0$ tel que

$$\frac{f\left(\hat{x}+\lambda x\right)-f\left(\hat{x}\right)}{\lambda^{2}} \geq 0, \quad \forall \lambda \in \left]-\delta,+\delta\right[.$$

Si on prend en considération (2) et on passe à la limite quand $\lambda \to 0$, $\lambda \neq 0$, on obtient

$$x^{t} H\left(\hat{x}\right) x \geq 0, \ \forall x \in \mathbb{R}^{n}.$$

■

4.5 Conditions suffisantes d'optimalité

Théorème Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois differentiable au point $\hat{x} \in \mathbb{R}^n$.Si $\nabla f\left(\hat{x}\right) = 0$ et $H\left(\hat{x}\right)$ est définie positive alors $\hat{x}$ est un minimum local strict de (P).

Preuve. f étant deux fois differentiable au point $\hat{x}$, on aura pour tout $x \in \mathbb{R}^n$

$$f(x)=f\left(\hat{x}\right)+\frac{1}{2}\left(x-\hat{x}\right)^{t} H\left(\hat{x}\right)\left(x-\hat{x}\right)+\left\|\left(x-\hat{x}\right)\right\|^{2} \alpha\left(\hat{x},\left(x-\hat{x}\right)\right), \quad \alpha\left(\hat{x},\left(x-\hat{x}\right)\right) \underset{x \to \hat{x}}{\longrightarrow} 0, \tag{3}$$

Supposons que $\hat{x}$ n'est pas un optimum local strict. Alors il existe une suite $\{x_k\}_{k \in \mathbb{N}\mathbb{N}^*}$ telle que $x_k \neq \hat{x} : \forall k$ et

$$x_k \neq \hat{x} : \forall k, \quad x_k \underset{k \to \infty}{\longrightarrow} \hat{x} \ \text{ et } \quad f(x_k) \leq f\left(\hat{x}\right). \tag{4}$$

Dans (3) prenons $x = x_k$, divisons le tout par $\left\|\left(x_k-\hat{x}\right)\right\|^{2}$ et notons $d_k = \frac{\left(x_k-\hat{x}\right)}{\left\|\left(x_k-\hat{x}\right)\right\|}$, on obtient

$$\frac{f(x_k)-f\left(\hat{x}\right)}{\left\|\left(x_k-\hat{x}\right)\right\|^{2}}=\frac{1}{2} d_k^{t} H\left(\hat{x}\right) d_k+\alpha\left(\hat{x},\left(x_k-\hat{x}\right)\right), \quad \alpha\left(\hat{x},\left(x_k-\hat{x}\right)\right) \underset{k \to \infty}{\longrightarrow} 0 \tag{5}$$

(4) et (5) impliquent

$$\frac{1}{2}d_k^t H\left(\hat{x}\right) d_k + \alpha\left(\hat{x}, \left(x_k - \hat{x}\right)\right) \leq 0, \qquad \forall k.$$

D'autre part la suite $\{d_k\}_{k\in\mathbb{N}\mathbb{N}^*}$ est bornée $(\|d_k\| = 1, \quad \forall k)$. Donc il existe une sous suite $\{d_k\}_{k\in N_1\subset\mathbb{N}\mathbb{N}}$ telle que

$$d_k \underset{k\to\infty,\ k\in N_1}{\longrightarrow} \tilde{d}.$$

Finalement lorsque $k \to \infty$, $k \in N_1$, on obtient

$$\frac{1}{2}\tilde{d}H\left(\hat{x}\right)\tilde{d} \ \leq 0.$$

La dernière relation et le fait que $\tilde{d} \neq 0$ $\left(\left\|\tilde{d}\right\| = 1\right)$ impliquent que la matrice hessienne $H\left(\hat{x}\right)$ n'est pas définie positive. Ceci est en contradiction avec l'hypothèse. ■

4.6 Les modes de convergence

Définition : Soit $\left\{x^k\right\}_{k\in\mathbb{N}}$ une suite dans $\mathbb{R}^n$ convergeant vers x^*.

1- Si $\lim\limits_{k\to\infty}\sup \frac{\left\|x^{k+1}-x^*\right\|}{\left\|x^k-x^*\right\|} = \alpha < 1$, on dit que $\left\{x^k\right\}_{k\in\mathbb{N}}$ converge vers x^* linéairement avec le taux α.

Lorsque $\left\|x^{k+1} - x^*\right\| \simeq \alpha\left\|x^k - x^*\right\|$, la convergence est dite linéaire asymptotique.

2- Si $\lim\limits_{k\to\infty}\sup \frac{\left\|x^{k+1}-x^*\right\|}{\left\|x^k-x^*\right\|^\gamma} < +\infty, \gamma > 1$ la convergence est dite superlinéaire d'ordre γ.

3- Si $\lim\limits_{k\to\infty} \frac{\left\|x^{k+1}-x^*\right\|}{\left\|x^k-x^*\right\|} = 0$, on dit que $\left\{x^k\right\}_{k\in\mathbb{N}}$ tend vers x^* de façon superlinéaire.

Lorsque $\left\|x^{k+1} - x^*\right\| \leq l\ \left\|x^k - x^*\right\|^\gamma$, la convergence est dite superlinéaire asymptotique.