Cours d'analyse numérique de licence de mathématiques

Roland Masson

2014

1 Optimisation sans contraintes

Optimisation sans contraintes: plan

- Notions de minimum local et global
- Conditions nécessaires et suffisantes de minimum local et global
- Algorithmes de descente

Notion de minimum local et global

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction, alors on dit que

- $\bar{x} \in \mathbb{R}^n$ est un minimum global de f ssi $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^n$
- $\bar{x} \in \mathbb{R}^n$ est un minimum local de f ssi il existe un voisinage U de $\bar{x}$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in U$.
- $\bar{x} \in \mathbb{R}^n$ est un minimum local strict de f ssi il existe un voisinage U de $\bar{x}$ tel que $f(\bar{x}) < f(x)$ pour tout $x \in U \setminus \{\bar{x}\}$.

Condition nécessaire d'ordre 1: équation d'Euler

Théorème: condition nécessaire d'ordre 1 de minimum local Soient $f \in C^1(\mathbb{R}^n, \mathbb{R})$ et $\bar{x} \in \mathbb{R}^n$ un minimum local de f. Alors nécessairement

$$\nabla f(\bar{x}) = 0 \text{ ie de façon équivalente } f'(\bar{x}) = 0.$$

Preuve: Supposons que $p = -\nabla f(\bar{x}) \in \mathbb{R}^n$ soit non nul, alors il existe $T > 0$ tel que pour tout $t \in [0, T]$ on ait par la formule des accroissements finis

$$f(\bar{x} + tp) = f(\bar{x}) + t\Big(\nabla f(\bar{x} + \theta(t)tp), p\Big)$$

avec $\theta(t) \in (0, 1)$, et par continuité de $\nabla f(x)$

$$\Big(\nabla f(\bar{x} + \theta(t)tp), p\Big) < 0,$$

ce qui contredit la condition de minimum local. On a donc bien $p = 0$.

Conditions nécessaires d'ordre 2

Théorème: Conditions nécessaires d'ordre 2 de minimum local Soient $f \in C^2(\mathbb{R}^n, \mathbb{R})$, et $\bar{x} \in \mathbb{R}^n$ un minimum local de f. Alors nécessairement

$$\nabla f(\bar{x}) = 0 \text{ et } H(\bar{x}) \text{ est positive.}$$

où $H(x)$ est la matrice Hessienne (représentant $f''(x)$).

Preuve: on sait déjà que $\nabla f(\bar{x}) = 0$ et que $H(\bar{x})$ est symétrique. Supposons qu'il existe $p \in \mathbb{R}^n$, $p \neq 0$ tel que

$$\Big(H(\bar{x})p, p\Big) < 0.$$

Alors, par la formule de Taylor à l'ordre 2 et par continuité de $H(x)$, il existe $T > 0$ tel que pour tout $t \in [0, T]$ on ait

$$f(\bar{x} + tp) = f(\bar{x}) + \frac{1}{2}t^2\Big(H(\bar{x} + t\theta(t)p)p, p\Big)$$

et $\Big(H(\bar{x} + t\theta(t)p)p, p\Big) < 0$, ce qui contredit la condition de minimum local. On déduit donc que $H(\bar{x})$ est positive.

Conditions suffisantes de minimum local strict

Théorème: condition suffisante de minimum local strict Soit $f \in C^2(\mathbb{R}^n, \mathbb{R})$ telle que $\nabla f(\bar{x}) = 0$ et que la matrice Hessienne $H(\bar{x})$ soit SDP, alors $\bar{x} \in \mathbb{R}^n$ est un minimum local strict de f.

On va tout d'abord démontrer le lemme suivant:
Lemme: soit $A(x) : \mathbb{R}^n \to \mathcal{S}_n(\mathbb{R})$ une application continue de $\mathbb{R}^n$ dans l'espace vectoriel des matrices symétriques, et soit $\bar{x} \in \mathbb{R}^n$ tel que $A(\bar{x})$ soit SDP. Alors il existe un voisinage V de $\bar{x}$ tel que $A(x)$ soit SDP pour tout $x \in V$.

Preuve du lemme: soit $\lambda_{min} > 0$ la valeur propre minimale de $A(\bar{x})$, on a

$$\inf_{\{z \in \mathbb{R}^n \mid \|z\|_2 = 1\}} (A(\bar{x})z, z) = \lambda_{min}.$$

Pour tous $z \in \mathbb{R}^n$ avec $\|z\|_2 = 1$ on a

$$(A(x)z, z) = (A(\bar{x})z, z) + (A(x) - A(\bar{x})z, z) \geq (A(\bar{x})z, z) - \|A(x) - A(\bar{x})\|_2,$$

et donc

$$\inf_{\{z \in \mathbb{R}^n \mid \|z\|_2 = 1\}} (A(x)z, z) \geq \lambda_{min} - \|A(x) - A(\bar{x})\|_2,$$

ce qui montre le lemme.

Conditions suffisantes de minimum local strict

Preuve du théorème: d'après la formule de Taylor à l'ordre 2 et le lemme précédent appliqué à $H(x)$, il existe $r > 0$ tel que pour tout $p \in \mathbb{R}^n$ avec $\|p\| < r$ on ait

$$f(\bar{x} + p) = f(\bar{x}) + \frac{1}{2}\Big(H(\bar{x} + \theta(p)p)p, p\Big)$$

avec $H(\bar{x} + \theta(p)p)$ SDP, ce qui démontre la condition de minimum local strict pour $\bar{x}$.

Minimum global et convexité

Fonction convexe: on dit que f de $\mathbb{R}^n$ dans $\mathbb{R}$ est convexe ssi pour tout $\theta \in [0,1]$

$$f(\theta y + (1-\theta)x) \leq \theta f(y) + (1-\theta) f(x),$$

pour tous $x, y \in \mathbb{R}^n \times \mathbb{R}^n$. f est dite strictement convexe si l'inégalité est stricte pour tout $\theta \in]0,1[$.

Théorème: Si f est convexe de $\mathbb{R}^n$ dans $\mathbb{R}$ alors tout minimum local de f est un minimum global de f.

Preuve: Soit $\bar{x}$ un minimum local de f. Si il existe $z \in \mathbb{R}^n$ tel que $f(z) < f(\bar{x})$, alors par convexité de f on a pour $\theta \in]0,1]$

$$f(\theta z + (1-\theta)\bar{x}) \leq \theta f(z) + (1-\theta) f(\bar{x}) < f(\bar{x}),$$

ce qui contredit la définition du minimum local.

Minimum global et convexité

Théorème: Soit f une fonction convexe de $\mathbb{R}^n$ dans $\mathbb{R}$. Si $\bar{x} \in \mathbb{R}^n$ vérifie l'équation d'Euler $\nabla f(\bar{x}) = 0$, alors $\bar{x}$ est un minimum global de f.

Preuve: Si il existe $z \in \mathbb{R}^n$ tel que $f(z) < f(\bar{x})$, alors soit $\phi(t) = f((1-t)\bar{x} + tz)$

$$\begin{array}{rcl} \Big(\nabla f(\bar{x}), (z - \bar{x})\Big) = \phi'(0) & = & lim_{t \to 0} \frac{f((1-t)\bar{x}+tz)-f(\bar{x})}{t} \\ & \leq & lim_{t \to 0} \frac{(1-t)f(\bar{x})+tf(z)-f(\bar{x})}{t} \\ & \leq & f(z) - f(\bar{x}) < 0, \end{array}$$

ce qui contredit $\nabla f(\bar{x}) = 0$.

Resultat d'existence et d'unicité d'un minimum global

Théorème: Soit $f \in C^0(\mathbb{R}^n, \mathbb{R})$ telle que $\lim_{\|x\| \to +\infty} = +\infty$, alors f admet un minimum global. Si de plus f est strictement convexe, alors ce minimum est unique.

Preuve: l'existence se déduit du fait qu'une fonction continue atteint ses bornes sur un compact. Ensuite si on a deux minima globaux distinct $\bar{x} \neq \bar{y}$ avec $m = f(\bar{x}) = f(\bar{y})$, alors

$$f(\frac{\bar{x} + \bar{y}}{2}) < m,$$

ce qui est une contradiction.

Algorithme de descente

Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$, on cherche à trouver un minimum local de f par un algorithme de descente suivant

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)},$$

où la direction de descente $p^{(k)}$ satisfait la condition

$$\left(\nabla f(x^{(k)}), p^{(k)}\right) < 0,$$

et $\alpha^{(k)} > 0$ est appelé le pas de descente.
On considèrera deux choix pour $p^{(k)}$:

- $p^{(k)} = -\nabla f(x^{(k)})$: algorithme de descente selon le gradient
- $p^{(k)} = -\left(H(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$: algorithme de Newton

Nb: l'algorithme de Newton définit une direction de descente si $\left(H(x^{(k)})\right)^{-1}$ est SDP.

Recherche linéaire

La recherche du pas $\alpha^{(k)}$ à direction de descente $p^{(k)}$ fixée s'appelle la recherche linéaire.
On cherche un pas vérifiant les **conditions de Wolfe** suivantes:

$$f(x^{(k)} + \alpha^{(k)} p^{(k)}) \leq f(x^{(k)}) + c_1 \alpha^{(k)} \Big(\nabla f(x^{(k)}), p^{(k)} \Big)$$
$$\Big(\nabla f(x^{(k)} + \alpha^{(k)} p^{(k)}), p^{(k)} \Big) \geq c_2 \Big(\nabla f(x^{(k)}), p^{(k)} \Big)$$

avec $0 < c_1 < c_2 < 1$.

Théorème: Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$ et soit p^k une direction de descente au point $x^{(k)}$. On suppose que $\phi(\alpha) = f(x^{(k)} + \alpha p^{(k)})$ est bornée inférieurement pour $\alpha > 0$, alors pour $0 < c_1 < c_2 < 1$, il existe $\alpha^{(k)} > 0$ satisfaisant les conditions de Wolfe.

Recherche linéaire

Preuve d'existence de $\alpha^{(k)}$ satisfaisant les conditions de Wolfe: Comme $\phi(\alpha)$ est bornée inférieurement et que $l(\alpha) = f(x^{(k)}) + c_1\alpha\Big(\nabla f(x^{(k)}), p^{(k)}\Big)$ tend vers $-\infty$ en $+\infty$, les deux graphes s'intersectent. Soit $\alpha' > 0$ la plus petite valeur de α avec intersection des deux graphes. On a donc

$$f(x^{(k)} + \alpha' p^{(k)}) = f(x^{(k)}) + c_1\alpha'\Big(\nabla f(x^{(k)}), p^{(k)}\Big)$$

Pour toutes les valeurs de $0 < \alpha < \alpha'$, le graphe de $\phi(\alpha)$ est au dessous de celui de $l(\alpha)$ et donc la première condition de Wolfe est vérifiée pour tout $0 < \alpha \leq \alpha'$. Ensuite, par le théorème des accroissements finis il existe $0 < \alpha'' < \alpha'$ tel que

$$f(x^{(k)} + \alpha' p^{(k)}) - f(x^{(k)}) = \alpha'\Big(\nabla f(x^{(k)} + \alpha'' p^{(k)}), p^{(k)}\Big) = c_1\alpha'\Big(\nabla f(x^{(k)}), p^{(k)}\Big)$$

et donc $\Big(\nabla f(x^{(k)} + \alpha'' p^{(k)}), p^{(k)}\Big) > c_2\Big(\nabla f(x^{(k)}), p^{(k)}\Big)$, ce qui achève de démontrer le théorème.

Recherche linéaire: modèle quadratique

Si $f \in C^2(\mathbb{R}^n, \mathbb{R})$ et si on a accès à la Hessienne $H(x)$ de f, pour améliorer la recherche linéaire, on utilise l'approximation quadratique de $f(x^{(k)} + p)$ obtenue par le développement de Taylor à l'ordre 2:

$$m_k(p) = f(x^{(k)}) + (\nabla f(x^{(k)}), p) + \frac{1}{2}(H(x^{(k)})p, p).$$

On initialise alors la recherche linéaire par

$$\alpha_{quad}^{(k)} = \mathrm{argmin}_{\alpha > 0} m_k(\alpha p^{(k)}).$$

Si $H(x^{(k)})$ est SDP, on obtient

$$\alpha_{quad}^{(k)} = \frac{(-\nabla f(x^{(k)}), p^{(k)})}{(H(x^{(k)})p^{(k)}, p^{(k)})},$$

ce qui donne bien sûr $\alpha_{quad}^{(k)} = 1$ pour la direction du Newton.

Algorithme de Backtracking

En pratique, on utilise souvent l'algorithme de recherche linéaire suivant (qui dispense de la seconde condition de Wolfe au sens où il évite de prendre des α trop petits):

Algorithme de Backtracking: Choisir $\bar{\alpha} > 0$, $\rho \in]0, 1[$, $c_1 \in]0, 1[$ et poser $\alpha = \bar{\alpha}$.

Tant que $f(x^{(k)} + \alpha p^{(k)}) > f(x^{(k)}) + c_1 \alpha (\nabla f(x^{(k)}), p^{(k)})$

$\alpha \leftarrow \rho\alpha$

Fin Tant que

$\alpha^{(k)} = \alpha$.

On partira typiquement de la valeur $\bar{\alpha}$ donnée par le modèle quadratique $\alpha^{(k)}_{quad}$.

Théorème de convergence de la méthode de descente avec recherche linéaire

Théorème de convergence: Soit $f \in C^2(\mathbb{R}^n, \mathbb{R})$ telle que $\lim_{\|x\| \to +\infty} f(x) = +\infty$. On suppose que

$$\cos(\theta^{(k)}) = \frac{(p^{(k)}, -\nabla f(x^{(k)}))}{\|p^{(k)}\| \|\nabla f(x^{(k)}\|} > 0,$$

et que $\alpha^{(k)}$ satisfait les conditions de Wolfe, alors l'algorithme de descente selon le gradient

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)},$$

vérifie

$$\sum_{k \geq 0} (\cos(\theta^{(k)}))^2 \|\nabla f(x^{(k)})\|^2 < +\infty.$$

Théorème de convergence de la méthode de descente selon le gradient avec recherche linéaire

Preuve du théorème de convergence Comme $\lim_{\|x\|\to+\infty} f(x) = +\infty$, la suite $x^{(k)}$ telle que $f(x^{(k)})$ reste nécessairement dans un borné B et il existe C telle que $\|\nabla f(y) - \nabla f(x)\| \leq C\|y - x\|$ pour tous $x, y \in B$.
D'après la seconde condition de Wolfe on a

$$\left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), p^{(k)}\right) \geq (1 - c_2)(-\nabla f(x^{(k)}), p^{(k)}),$$

et par ailleurs $\left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), p^{(k)}\right) \leq \alpha^{(k)} C\|p^{(k)}\|^2$, ce qui implique que

$$\alpha^{(k)} \geq \frac{1 - c_2}{C} \frac{(-\nabla f(x^{(k)}), p^{(k)})}{\|p^{(k)}\|^2}.$$

En injectant cette relation dans la première condition de Wolfe on a donc que

$$f(x^{(k+1)}) \leq f(x^{(k)}) - c_1 \frac{1 - c_2}{C} (\cos(\theta^{(k)}))^2 \|\nabla f(x^{(k)})\|^2,$$

et donc que $f(x^{(k+1)}) \leq f(x^{(0)}) - c_1 \frac{1-c_2}{C} \sum_{q=0}^{k} (\cos(\theta^{(q)}))^2 \|\nabla f(x^{(q)})\|^2$.
Comme f est bornée inférieurement, ceci démontre le théorème.

Théorème de convergence de la méthode de descente selon le gradient: exemples

La méthode de descente selon le gradient vérifie les hypothèses du théorème car $\cos(\theta^{(k)}) = 1$. On en déduit que $\nabla f(x^{(k)})$ converge vers 0 et que $x^{(k)}$ converge donc vers un point stationnaire à une sous suite près.

En ce qui concerne l'algorithme de Newton, si on suppose que $H(x^{(k)})$ est SDP et que le conditionnement de $H(x^{(k)})$ est uniformément borné par K, alors on a

$$\cos(\theta^{(k)}) \geq \frac{1}{\mathrm{Cond}(H(x^{(k)}))} \geq \frac{1}{K}$$

et donc $\nabla f(x^{(k)})$ converge vers 0.

Algorithme de descente selon le gradient dans le cas quadratique

Soient $A \in \mathcal{M}_n(\mathbb{R})$ SDP et $b \in \mathbb{R}^n$, $c \in \mathbb{R}$, on considère

$$f(x) = \frac{1}{2}(Ax, x) - (b, x) + c,$$

alors A admet un minimum global unique tel que $\nabla f(x) = Ax - b = 0$.
Dans ce cas, l'algorithme de descente selon le gradient équivaut à l'algorithme de Richardson appliqué au système linéaire $Ax = b$ et on sait comment choisir $\alpha^{(k)}$ de façon à garantir la convergence: $r^{(k)} = -\nabla f(x^{(k)}) = b - Ax^{(k)}$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} r^{(k)},$$

$$\alpha^{(k)} = \frac{(r^{(k)}, r^{(k)})}{(Ar^{(k)}, r^{(k)})},$$

et on rappelle que l'on obtient un taux de convergence linéaire donné par

$$\|x^{(k+1)} - \bar{x}\|_A \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right) \|x^{(k)} - \bar{x}\ \|_A,$$

où $\|x\|_A = (Ax, x)^{1/2}$. La vitesse de convergence est donc dépendante du conditionnement de A.

Théorème de convergence local de l'algorithme de Newton

Théorème: Soit $f \in C^3(\mathbb{R}^n, \mathbb{R})$ et $\bar{x}$ tel que $\nabla f(\bar{x}) = 0$ et $H(\bar{x})$ soit SDP, alors il existe un voisinage V de $\bar{x}$ tel que si $x^{(0)} \in V$, l'algorithme de Newton

$$x^{(k+1)} = x^{(k)} - \left(H(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$$

converge au moins quadratiquement vers $\bar{x}$ qui est un minimum local strict de f.

Preuve: il s'agit de l'algorithme de Newton appliqué à l'équation d'Euler $\nabla f(x) = 0$. Le théorème est donc déjà connu. D'autre part on a aussi déjà montré que $\bar{x}$ est un minimum local strict de f.