le cnam

Analyse Mathématique pour l'Ingénieur

Paris, 2009 - 2016

Cours 08

Introduction à l'optimisation

François Dubois

le cnam

Analyse Mathématique pour l'Ingénieur

Cours 8
Introduction à l'optimisation *

- Formules de Taylor
- Conditions nécessaires pour un minimum
- Multiplicateurs de Lagrange

* François Dubois, 2013, édition septembre 2015, 11 pages.

Introduction à l'optimisation

1 Formules de Taylor.

- On commence par du calcul différentiel élémentaire pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$. Soit $\varphi : [0,1] \to \mathbb{R}$ aussi régulière que de besoin. On a :

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,dt = \Big[(t-1)\varphi'(t)\Big]_0^1 + \int_0^1 (1-t)\,\varphi''(t)\,dt$$

$$(1) \quad \varphi(1) = \varphi(0) + \varphi'(0) + \int_0^1 (1-t)\,\varphi''(t)\,dt.$$

- On peut intégrer une nouvelle fois le membre de droite de (1). par parties. Il vient

$$\int_0^1 (1-t)\,\varphi''(t)\,dt = -\left[\frac{(t-1)^2}{2}\varphi''(t)\right]_0^1 + \int_0^1 \frac{(1-t)^2}{2}\varphi^{(3)}(t)\,dt$$

$$(2) \quad \varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(0) + \int_0^1 \frac{(1-t)^2}{2}\varphi^{(3)}(t)\,dt.$$

- Soit $f : E \to \mathbb{R}$ où E est un espace de Banach, et $h \in E$ fixé. On étudie f au voisinage de $x \in E$ fixé. On pose

$$(3) \qquad \varphi(t) = f(x + th), \quad 0 \leq t \leq 1.$$

Alors (4) $\quad \varphi'(t) = df(x+th)\cdot h, \quad 0 \leq t \leq 1$

(5) $\varphi''(t) = d^2f(x+th)\cdot(h,h), \quad 0 \leq t \leq 1$

(6) $\varphi^{(3)}(t) = d^3f(x+th)\cdot(h,h,h), \quad 0 \leq t \leq 1$

par dérivation successive. En particulier,

$\varphi(1) = f(x+h)$, $\varphi(0) = f(x)$, $\varphi'(0) = df(x)\cdot h$, $\varphi''(0) = d^2f(x)\cdot(h,h)$. Les relations (1) et (2) permettent d'écrire les formules de Taylor avec reste intégral.

Prop ① <u>Formule de Taylor à l'ordre deux</u>

Si $f : E \to \mathbb{R}$ est deux fois continûment différentiable au voisinage de $x \in E$, on a

$$(7) \begin{cases} f(x+h) = f(x) + df(x)\cdot h + \\ \qquad + \int_0^1 (1-t)\, d^2f(x+th)\cdot(h,h)\, dt \end{cases}$$

Prop ② <u>Formule de Taylor à l'ordre trois.</u>

Si $f : E \to \mathbb{R}$ est trois fois continûment différentiable au voisinage de $x \in E$, on a

$$(8) \begin{cases} f(x+h) = f(x) + df(x)\cdot h + \frac{1}{2} d^2f(x)\cdot(h,h) \\ \qquad + \frac{1}{2}\int_0^1 (1-t)^2\, d^3f(x+th)\cdot(h,h,h)\, dt \end{cases}$$

② Conditions nécessaires de minimum.

Soit E un espace de Banach, f défini au voisinage de $a \in E$ (f est une application $B(a,r) \to E$ avec $r>0$ bien choisi) de sorte que f est minimale au point a:

(9) $f(x) \geq f(a), \quad \forall x \in B(a,r)$

Prop ③ Condition d'ordre un d'extremum

Avec les hypothèses précédentes, si f est différentiable en a, alors on a

(10) $df(a) = 0$

Preuve de la proposition ③

Le plus simple est de réintroduire, pour $h \in E$ fixé, la fonction $\varphi : [0,\alpha] \to \mathbb{R}$ par la relation (3). Le nombre α est tel que $\alpha \|h\| < r$. On a donc $\varphi(t) = f(a+th)$ pour $0 \leq t \leq \alpha$. Alors $\varphi(t) \geq \varphi(0)$ pour $0<t<\alpha$ (cf (9)) et $\frac{1}{t}(\varphi(t)-\varphi(0)) \geq 0$ dans les mêmes conditions. On fait tendre t vers 0. La fonction φ est dérivable en 0 car f, différentiable en a, est dérivable le long de tout vecteur h. On en déduit

$df(a)\cdot h = \varphi'(0) \geqslant 0$. Mais on peut recommencer le raisonnement avec $\tilde{\varphi}(t)$ obtenue en explorant la direction $-h$: $\tilde{\varphi}(t) = \varphi(a - th)$. De même que plus haut, $\frac{1}{t}(\tilde{\varphi}(t) - \tilde{\varphi}(0)) \geqslant 0$ quel que soit $t \in]0,\alpha]$.
En prenant la limite de cette expression pour t tendant vers 0, $df(a)\cdot(-h) = \tilde{\varphi}'(0) \geqslant 0$ et $df(a)\cdot h \leq 0$.
On en déduit que, pour tout $h \neq 0$ de E $df(a)\cdot h = 0$. D'où la nullité de l'application linéaire $df(a)$, et la relation (10) est établie.

$\square$

Prop (4) Condition nécessaire d'ordre deux.

Soit $f : B(a,r) \to \mathbb{R}$ deux fois continuement différentiable au voisinage de $a \in E$. On suppose f minimale en a ; l'inégalité (9) a lieu pour tout $x \in B(a,r)$. Alors pour $h \in E$ arbitraire, on a l'inégalité

(11) $\quad d^2f(a)\cdot(h,h) \geqslant 0$;

- la matrice Hessienne $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1 \leq i,j \leq N}$ des dérivées partielles secondes (dans le cas $E = \mathbb{R}^N$) est symétrique positive.

Preuve de la proposition (4).

On se donne t de sorte que $|t| \leq \alpha = \frac{r}{\|h\|}$ comme pour la proposition 3. On introduit $\varphi(t) \equiv f(x+th)$, défini sur l'intervalle $[-\alpha, \alpha]$ pour fixer les idées. On a alors le développement de Taylor, suivant, variante de (1) :

$$\varphi(t) = \varphi(0) + \int_0^t \varphi'(\theta)\, d\theta = \varphi(0) + \int_0^1 \varphi'(t\theta)\, t\, d\theta$$

$$= \varphi(0) + t \left[\left[(\theta-1)\, \varphi'(t\theta)\right]_0^1 + \int_0^1 (1-\theta)\, \varphi''(t\theta)\, t\, d\theta \right]$$

$$= \varphi(0) + t\,\varphi'(0) + t^2 \int_0^1 (1-\theta)\, \varphi''(t\theta)\, d\theta$$

$$(12) \quad \varphi(t) = \varphi(0) + t\varphi'(0) + t^2 \int_0^1 (1-\theta)\, \varphi''(t\theta)\, d\theta .$$

* Si (9) a lieu, $\varphi'(0) = 0$ (voir la proposition (3)) et pour $t \in [-\alpha, \alpha] \setminus \{0\}$, on a $\frac{1}{t^2}(\varphi(t) - \varphi(0)) \geq 0$, c'est à dire $\int_0^1 (1-\theta)\, \varphi''(t\theta)\, d\theta \geq 0$.
Si f est de classe $\mathcal{C}^2$ dans $B(a,r)$, alors $\varphi''(.)$ est continue et si t tend vers zéro, l'intégrale $\int_0^1 (1-\theta)\, \varphi''(t\theta)\, d\theta$ converge vers $\frac{1}{2}\varphi''(0)$.
Donc $\varphi''(0) \geq 0$. Or $\varphi''(0) = d^2 f(a) \cdot (h, h)$, ce qui établit l'inégalité (11). $\square$

3 Multiplicateurs de Lagrange

On se place dans le cas où $E = \mathbb{R}^N$ et on étudie une fonction $f : E \to \mathbb{R}$ sur l'ensemble

(13) $K = \{x \in \mathbb{R}^N,\ g(x) = 0\}$

où $g : \mathbb{R}^N \to \mathbb{R}$ est une fonction régulière (de classe au moins C^1 typiquement). On suppose que $K \neq \phi$ et on suppose que f admet un minimum (local) au point $a \in K$;

(14) $f(x) \geqslant f(a), \quad \forall x \in B(a,r) \cap K$.

Prop 5 Existence du multiplicateur.

Dans le cadre précédent, on suppose de plus que le point a est régulier :

(15) $g(a) = 0, \ \exists j \in \{1, \dots, N\}, \ \frac{\partial g}{\partial x_j}(a) \neq 0$.

alors il existe $\lambda \in \mathbb{R}$ tel que

(16) $df(a) - \lambda\, dg(a) = 0$

Le réel λ est le multiplicateur de Lagrange associé à la contrainte g

Preuve de la proposition (7)

* Dans la condition (15), on peut supposer pour fixer les idées que $\frac{\partial g}{\partial x_N}(a) \neq 0$ [i.e $j = N$ convient]. On peut appliquer au voisinage de a le théorème des fonctions implicites. Il existe un voisinage V ouvert de $a' = (a_1, \dots, a_{N-1})$ tel que, quitte à réduire $r > 0$ de la condition (14), la résolution $g(x) = 0$ avec $x \in B(a, r)$ s'explicite sous la forme $x_N = \varphi(x_1, \dots, x_{N-1})$, où $\varphi : V \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$.
La relation (14) peut donc s'écrire $f(x_1, \dots, x_{N-1}, \varphi(x_1, \dots, x_{N-1})) \geq f(a)$ pour tout $(x_1, \dots, x_{N-1})$ au voisinage de a'. La relation (10) appliquée à cette fonction entraîne

$$(17) \qquad \frac{\partial f}{\partial x_j}(a) + \frac{\partial f}{\partial x_N}(a) \cdot \frac{\partial \varphi}{\partial x_j}(a') = 0, \quad 1 \leq j \leq N-1.$$

* Par ailleurs, la définition (13) de K entraîne

$$(18) \qquad \frac{\partial g}{\partial x_j}(a) + \frac{\partial g}{\partial x_N}(a) \frac{\partial \varphi}{\partial x_j}(a') = 0, \quad 1 \leq j \leq N-1$$

Comme $\frac{\partial g}{\partial x_N}(a) \neq 0$ par hypothèse, on peut substituer $\frac{\partial \varphi}{\partial x_j}(a')$ tiré de (18) au sein de (17) :

$\frac{\partial \varphi}{\partial x_j}(a') = -\frac{\partial g}{\partial x_j}(a) \Big/ \frac{\partial g}{\partial x_N}(a)$ et on en déduit

$$(19) \quad \frac{\partial f}{\partial x_j}(a) - \frac{\frac{\partial f}{\partial x_N}(a)}{\frac{\partial g}{\partial x_N}(a)} \frac{\partial g}{\partial x_j}(a) = 0, \quad 1 \leq j \leq N-1.$$

Or la relation (19) est encore valable pour $j = N$. Quitte à poser $\lambda = -\partial f/\partial x_N(a) \Big/ \frac{\partial g}{\partial x_N}(a)$, on en déduit

$$(20) \quad \frac{\partial f}{\partial x_j}(a) - \lambda \frac{\partial g}{\partial x_j}(a) = 0, \quad 1 \leq j \leq N,$$

relation qui exprime exactement la relation (16). $\square$

- Par exemple, on peut chercher les points extremaux de $f(x,y) \equiv x^2 - y^2$ (on a ici $N=2$) sur le cercle $K = \{(x,y) \in \mathbb{R}^2, \; x^2+y^2 = 1\}$. On forme le Lagrangien

$$(21) \quad \mathcal{L}(x,y,\lambda) = f(x,y) - \lambda\, g(x,y).$$

La relation (16) s'écrit ici pour les deux dérivées partielles :

$$\frac{\partial \mathcal{L}}{\partial x} \equiv 2x - \lambda(2x) = 0, \quad \frac{\partial \mathcal{L}}{\partial y} \equiv -2y - 2\lambda y = 0.$$

D'où $x(1-\lambda) = 0$ et $y(1+\lambda) = 0$.

* Si $x = 0$, alors $y = \pm 1$ puisque $(x,y) \in K$. Alors $\lambda = -1$ et on a $(\nabla f + \nabla g)(0, \pm 1) = 0$ qui exprime que le gradient de la fonction à minimiser et le gradient de la contrainte

sont non nuls et proportionnels.

* Si $x \neq 0$, alors $\lambda = 1$ et $(1+\lambda) \neq 0$. Donc $y = 0$ et $x = \pm 1$ car $(x,y) \in K$. On a donc $(\nabla f - \nabla g)(\pm 1, 0) = 0$; la proportionnalité des deux gradients prend une autre forme en ces deux nouveaux points extrémaux.
* Il importe ensuite de déterminer si on est en présence d'un maximum ou d'un minimum. Cet exercice est laissé au lecteur.

• Il importe dans les applications de savoir interpréter le multiplicateur de Lagrange en le reliant aux données du problème. C'est le cas si $x \in \mathbb{R}^N$ satisfait à la condition $x \in K_\xi$:

$$(22) \qquad K_\xi = \{x \in \mathbb{R}^N,\ g(x) = \xi\},$$

paramétrée par $\xi \in \mathbb{R}^p$, avec $g : \mathbb{R}^N \to \mathbb{R}^p$ pour enrichir un peu le problème. On cherche à minimiser $f : \mathbb{R}^N \to \mathbb{R}$ sous la contrainte $x \in K_\xi$. Sous l'hypothèse de régularité au point de minimum a_ξ, on a la condition (20) qui devient

$$(23) \qquad \frac{\partial f}{\partial x_j}(a_\xi) - \sum_{k=1}^{p} \lambda_k \frac{\partial g_k}{\partial x_j}(a) = 0, \quad 1 \leq j \leq N$$

on a le résultat suivant :

Prop (6) Interprétation du multiplicateur.

Dans le cadre défini plus haut, le gradient $\frac{\partial}{\partial \xi_k} f(a_\xi)$ de la valeur minimale de la fonction f sur l'ensemble K_ξ défini en (22) est égal au multiplicateur de Lagrange λ_k de la condition (23).

Preuve de la proposition (6)

- on introduit le Lagrangien, qui dépend maintenant de $x \in \mathbb{R}^N$, de $\lambda \in \mathbb{R}^P$ et de $\xi \in \mathbb{R}^P$:

(24) $\mathcal{L}(x,\lambda,\xi) = f(x) - (\lambda, g(x) - \xi)$.

on a le calcul classique $d\mathcal{L}(x,\lambda,\xi) \bullet (h,\mu,\zeta) =$

$$\sum_{j=1}^{N} \frac{\partial \mathcal{L}}{\partial x_j} h_j + \sum_{k=1}^{P} \frac{\partial \mathcal{L}}{\partial \lambda_k} \mu_k + \sum_{k=1}^{P} \frac{\partial \mathcal{L}}{\partial \xi_k} \zeta_k =$$

$$= \sum_{j=1}^{N} \Big[\frac{\partial f}{\partial x_j} - \sum_k \lambda_k \frac{\partial g_k}{\partial x_j} \Big] h_j - \sum_{k=1}^{P} (g_k(x) - \xi_k) \mu_k$$

$$+ \sum_{k=1}^{P} \lambda_k \zeta_k .$$ Au point de minimum a_ξ, on a $g(a_\xi) - \xi = 0$, donc $\mathcal{L}(a_\xi, \lambda, \xi) = f(a_\xi)$. Par ailleurs, $\big(\partial_j f - \sum_k \lambda_k \partial_j f\big)(a_\xi)$ et $g_k(a_\xi) - \xi$ sont nuls. Donc

(25) $$df(a_\xi) \bullet \zeta = \sum_{k=1}^{P} \lambda_k \xi_k ,$$

ce qui montre la propriété. $\square$

- Pour l'exemple dans $\mathbb{R}^2$ considéré plus haut, on peut changer le cercle unité en $K_\xi = \{(x,y)\in\mathbb{R}^2,\ x^2+y^2=\xi\}$. Les extréma sont maintenant aux quatre points $(\pm\sqrt{\xi},0)$ et $(0,\pm\sqrt{\xi})$. Dans le premier cas, le multiplicateur de Lagrange vaut $+1$ et il vaut -1 dans le second cas. La valeur (maximale) $f(\pm\sqrt{\xi},0)$ vaut ξ dans le premier cas et $f(0,\pm\sqrt{\xi}) = -\xi$ dans le second cas. On a donc clairement $\frac{\partial}{\partial\xi} f(\pm\sqrt{\xi},0) = +1$ et $\frac{\partial}{\partial\xi} f(0,\pm\sqrt{\xi}) = -1$ dans chaque cas; cet exemple explicite la relation (25) dans ce cas particulier.

Dubois

Paris, 24 novembre 2013.