Théorème des extrema liés [1, 2]

Leçons : 151, 159, 214, 215, 219

[Gou An], partie 5.3.2

Théorème

Soient $f, g_1, \ldots, g_r : U \to \mathbb{R}$ des fonctions de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$.
On pose $\Gamma = \{x \in U | \forall i \in [\![1, r]\!], g_i(x) = 0\}$.
On suppose que :
- $f|_\Gamma$ admet un extremum local en $a \in \Gamma$;
- les formes linéaires $\mathrm{D}g_1(a), \ldots, \mathrm{D}g_r(a)$ sont linéairement indépendantes.

Alors il existe $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$, appelés multiplicateurs de Lagrange, tels que : $\mathrm{D}f(a) = \sum_{i=1}^{r} \lambda_i \mathrm{D}g_i(a)$.

1. On peut aussi écrire "extremums" ou "extrémums", mais pas "extréma", "extremas" ou "extrémas".
2. Donnons de ce théorème quelques applications.

Le théorème spectral : Soit E un espace euclidien, $u \in \mathcal{L}(E)$ un endomorphisme symétrique (c'est-à-dire tel que $u^* = u$) ; alors il existe une base orthonormée de E formée de vecteurs propres de u. On considère les applications différentiables :

$$f : \left| \begin{array}{ccc} E & \to & \mathbb{R} \\ x & \mapsto & \langle u(x), x \rangle \end{array} \right. \quad \text{et } g : \left| \begin{array}{ccc} E & \to & \mathbb{R} \\ x & \mapsto & \langle x, x \rangle \end{array} \right. .$$

On note aussi $\mathcal{S} = \{x \in E | g(x) = 1\}$ la sphère unité ; on est en dimension finie, donc elle est compacte. L'application f étant continue sur E, elle atteint son maximum sur $\mathcal{S}$, en un point noté e_1. Par ailleurs, pour $x \in E$ et $h \in E$:

$$\mathrm{D}f(x).h = 2\langle u(x), h \rangle \text{ et } \mathrm{D}g(x).h = 2\langle x, h \rangle.$$

D'après le théorème des extrema liés, $\exists \lambda_1 \in \mathbb{R}, \mathrm{D}f(e_1) = \lambda_1 \mathrm{D}g(e_1)$. Autrement dit, $u(e_1) = \lambda_1 e_1$. On peut à présent raisonner par récurrence ; soit $F = e_1^\perp$. Il suffit de montrer que $u_{|F}$ est symétrique et appliquer la récurrence.

L'inégalité entre les moyennes arithmétique et géométrique : On considère les applications :

$$f : \left| \begin{array}{ccc} (\mathbb{R}^+)^n & \to & \mathbb{R} \\ x & \mapsto & \sqrt[n]{x_1 \ldots x_n} \end{array} \right. \quad \text{et } g : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \frac{1}{n}\sum_{i=1}^n x_i - 1 \end{array} \right. .$$

L'ensemble $K = \{x \in \mathbb{R}^n | x_1 \geqslant 0, \ldots x_n \geqslant 0, g(x) = 0\}$ est un compact de $\mathbb{R}^n$. La fonction f atteint son maximum sur K en un point $a = (a_1, \ldots, a_n)$; notamment $f(a) \geqslant f(1, \ldots, 1) = 1$. En conséquence, $a \in \left(\mathbb{R}^{+*}\right)^n$, ouvert sur lequel f et g sont de classe $\mathcal{C}^1$. Par ailleurs, pour $x \in \left(\mathbb{R}^{+*}\right)^n$, $\mathrm{D}g(x) = \dfrac{1}{n}(1, \ldots, 1)$. D'après le théorème des extrema liés, $\exists \lambda \in \mathbb{R}, \mathrm{D}f(a) = \lambda \mathrm{D}g(a)$. On montre alors que $\dfrac{\partial f}{\partial x_i}(a) = \dfrac{f(a)}{n a_i}$ pour $i \in [\![1, n]\!]$. En conséquence, $f(a) = \lambda a_1 = \ldots = \lambda a_n$, puis $a_1 = \ldots = a_n = 1$. Ainsi, $\forall x \in K, f(x) \leqslant 1$; puis, par homogénéité : $\forall x_1, \ldots, x_n \in \mathbb{R}^+$, $\sqrt[n]{x_1 \ldots x_n} \leqslant \dfrac{x_1 + \ldots + x_n}{n}$.

Une utilisation en statistiques : Soit $X_1, \ldots, X_n$ un n-échantillon de moyenne ν et de variance σ^2.

Quand $\sum_{i=1}^n a_i = 1$, on sait que $\sum_{i=1}^n a_i X_i$ est un estimateur non-biaisé de ν. Parmi tous les estimateurs de cette forme, on recherche celui de variance minimale.
Par le calcul, la variance de l'estimateur $\sum_{i=1}^n a_i X_i$ vaut $\sigma^2 \sum_{i=1}^n a_i^2$.
On pose :

$$f : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & x_1^2 + \ldots + x_n^2 \end{array} \right. \quad \text{et } g : \left| \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & x_1 + \ldots + x_n - 1 \end{array} \right. .$$

On veut minimiser f sur l'ensemble où g s'annule.
Par le théorème des extrema liés, si f admet un extremum en $(a_1, \ldots, a_n)$, alors

$$\exists \lambda \in \mathbb{R}, \mathrm{D}f(a_1, \ldots, a_n) = \lambda \mathrm{D}g(a_1, \ldots, a_n).$$

Mais $\dfrac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = 2a_i$ et $\dfrac{\partial g}{\partial x_i}(a_1, \ldots, a_n) = 1$.
Donc $\forall i \in [\![1, n]\!], 2a_i = 1$.
On montre alors que $a_1 = \ldots = a_n = \dfrac{1}{n}$; c'est bien un minimum, car $f(1, 0, \ldots, 0) = 1 > n\dfrac{1}{n^2} = \dfrac{1}{n}$.

Démonstration :

Étape 1 : Commençons par quelques petites remarques.

On voit déjà que nécessairement $r \leqslant n$ car les formes linéaires $\mathrm{D}g_1(a), \ldots, \mathrm{D}g_r(a)$ forment une famille libre de $(\mathbb{R}^n)^*$.

De plus, si $r = n$, le résultat est trivial, car alors $\mathrm{D}g_1(a), \ldots, \mathrm{D}g_r(a)$ forment une base de $(\mathbb{R}^n)^*$.

On peut donc supposer désormais que $r < n$.

Soit $s = n - r \geqslant 1$; on procède à l'identification entre $(x, y) \in \mathbb{R}^s \times \mathbb{R}^r$ et $(x_1, \ldots, x_s, y_1, \ldots, y_r) \in \mathbb{R}^n$.

On écrit alors $a = (\alpha, \beta)$, où $\alpha \in \mathbb{R}^s$ et $\beta \in \mathbb{R}^r$.

Étape 2 : Intéressons-nous à une matrice qui va nous être utile pour la suite.[3]

Soit $A = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_s} & \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_r} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_r}{\partial x_1} & \cdots & \frac{\partial g_r}{\partial x_s} & \frac{\partial g_r}{\partial y_1} & \cdots & \frac{\partial g_r}{\partial y_r} \end{pmatrix}(a) \in \mathcal{M}_{r,n}(\mathbb{R})$.[4]

Comme $(\mathrm{D}g_i(a))_{1 \leqslant i \leqslant r}$ est une famille libre, on a : $\operatorname{rg} A = r$.[5]

On peut donc extraire de A une sous-matrice inversible de format $r \times r$; quitte à renuméroter les variables... on peut supposer que $\det \left(\frac{\partial g_i}{\partial y_j}(a) \right)_{1 \leqslant i,j \leqslant r} \neq 0$!

En notant $g = (g_1, \ldots, g_r)$, ceci se reformule en "$\mathrm{D}_y g(a)$ est inversible".

Étape 3 : On applique le théorème des fonctions implicites à g au voisinage de a.[6]

Il nous fournit ici :

- U', voisinage ouvert de α dans $\mathbb{R}^s$,
- Ω, voisinage ouvert de a dans $\mathbb{R}^n$ et
- $\varphi = (\varphi_1, \ldots, \varphi_r) : U' \to \mathbb{R}^r$ de classe $\mathcal{C}^1$, tels que :

$$(x \in U', (x, y) \in \Omega \text{ et } g(x, y) = 0) \Leftrightarrow (x \in U' \text{ et } y = \varphi(x)).$$

En d'autres termes, les éléments de $\Gamma \cap \Omega$ s'écrivent $(x, \varphi(x))$.

Étape 4 : On pose $h : \left| \begin{array}{ccc} U' & \to & \mathbb{R} \\ x & \mapsto & f(x, \varphi(x)) \end{array} \right.$.

Comme $h(\alpha) = f(a)$ et $\forall x \in U', (x, \varphi(x)) \in \Gamma$, ensemble où f admet un extremum local en a ; on obtient que h admet un extremum local en α.

On note $\psi = (\mathrm{Id}_{\mathbb{R}^s}, \varphi)$.

---

3. Si elle ne servait à rien, on n'en parlerait pas...
4. Cette notation, bâtarde, signifie

$$A = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_s}(a) & \frac{\partial g_1}{\partial y_1}(a) & \cdots & \frac{\partial g_1}{\partial y_r}(a) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_r}{\partial x_1}(a) & \cdots & \frac{\partial g_r}{\partial x_s}(a) & \frac{\partial g_r}{\partial y_1}(a) & \cdots & \frac{\partial g_r}{\partial y_r}(a) \end{pmatrix}.$$

Elle a le mérite de prendre moins de place au tableau.

5. Cela se démontre par l'absurde ; supposons que $\operatorname{rg} A < r$. Alors il existe une famille de scalaires non-nulle $(\mu_1, \ldots, \mu_r)$ telle que : $\forall j \in [\![1, n]\!], \sum_{i=1}^{r} \mu_i \frac{\partial g_i}{\partial z_j}(a) = 0$, où on a désigné les variables $(x_1, \ldots, x_s, y_1, \ldots, y_r)$ par $(z_1, \ldots, z_n)$. En conséquence, $\sum_{i=1}^{r} \mu_i \mathrm{D}g_i(a) = 0$; contredisant ainsi la liberté de la famille $(\mathrm{D}g_i(a))_{1 \leqslant i \leqslant r}$.
6. Si je rappelle son énoncé, c'est pas pour vous offenser, c'est juste que ça me fait du bien.

Théorème (*des fonctions implicites*)

Soient U un ouvert de $\mathbb{R}^s \times \mathbb{R}^r$, (a, b) un point de U, et $f \in \mathcal{C}^1(U, \mathbb{R}^r)$.

On suppose que $f(a, b) = 0$ et que $\det \mathrm{D}_y f(a, b) \neq 0$.

Alors l'équation $f(x, y) = 0$ peut être résolue localement par rapport aux variables y, c'est-à-dire : il existe un voisinage ouvert V de a dans $\mathbb{R}^s$, un voisinage ouvert W de b dans $\mathbb{R}^r$, avec $V \times W \subset U$ et une unique application $\varphi : V \to W$ de classe $\mathcal{C}^1$ telle que :

$$(x \in V, y \in W \text{ et } f(x, y) = 0) \Leftrightarrow (x \in V \text{ et } y = \varphi(x)).$$

De plus, $\mathrm{D}_y f(x_0, y_0)$ est inversible pour tout $(x_0, y_0) \in V \times W$.

Ainsi : $\forall i \in [\![1,s]\!], 0 = \dfrac{\partial h}{\partial x_i}(\alpha) = \dfrac{\partial (f \circ \psi)}{\partial x_i}(\alpha) = \displaystyle\sum_{j=1}^{s} \dfrac{\partial f}{\partial x_j}(\psi(\alpha))\dfrac{\partial \psi_j}{\partial x_i}(\alpha) + \sum_{j=1}^{r} \dfrac{\partial f}{\partial y_j}(\psi(\alpha))\dfrac{\partial \psi_{s+j}}{\partial x_i}(\alpha)$.

Cependant, $\forall j \in [\![1,s]\!], \dfrac{\partial \psi_j}{\partial x_i} = \delta_{i,j}$ et $\forall j \in [\![1,r]\!], \dfrac{\partial \psi_{s+j}}{\partial x_i} = \dfrac{\partial \varphi_j}{\partial x_i}$.

Dès lors : $\forall i \in [\![1,s]\!], 0 = \dfrac{\partial f}{\partial x_i}(a) + \displaystyle\sum_{j=1}^{r} \dfrac{\partial f}{\partial y_j}(a)\dfrac{\partial \varphi_j}{\partial x_i}(\alpha)$.

De plus, $g \circ \psi$ est nulle sur U' donc sa $k^{\text{ème}}$ composante, $g_k \circ \psi$ (où $k \in [\![1,r]\!]$), aussi ; en conséquence :
$\forall i \in [\![1,s]\!], 0 = \dfrac{\partial g_k}{\partial x_i}(a) + \displaystyle\sum_{j=1}^{r} \dfrac{\partial g_k}{\partial y_i}(a)\dfrac{\partial \varphi_j}{\partial x_i}(\alpha)$.

Posons alors :

$$M = \left(\begin{array}{cccccc} \frac{\partial f}{\partial x_1}(a) & \dots & \frac{\partial f}{\partial x_s}(a) & \frac{\partial f}{\partial y_1}(a) & \dots & \frac{\partial f}{\partial y_r}(a) \\ \hline \\ & & & A & & \end{array}\right) \in \mathcal{M}_{r+1,n}(\mathbb{R}).$$

Les s premières colonnes de M sont combinaisons linéaires des r dernières ; ainsi rg $M \leqslant r$.
Les $r+1$ lignes de M sont alors liées !

Ainsi, $\exists\,(\mu_0, \dots, \mu_r) \in \mathbb{R}^{r+1}\backslash\{0\}, \mu_0 \mathrm{D}f(a) + \displaystyle\sum_{i=1}^{r} \mu_i \mathrm{D}g_i(a) = 0$.

Mais $(\mathrm{D}g_i(a))_{1 \leqslant i \leqslant r}$ est une famille libre, donc $\mu_0 \neq 0$ (car sinon, tous les μ_i devraient être nuls).

On obtient alors le résultat souhaité en posant : $\forall i \in [\![0,r]\!], \lambda_i = \dfrac{\mu_i}{\mu_0}$. ■

Références

[Gou An] X. GOURDON – *Les maths en tête : Analyse*, 2e éd., Ellipses, 2008.