
Optimisation

Université Paris Cité 2025-2026

M1 MMAS

Quentin Denoyelle & Camille Pouchol

1	Rappels de calcul différentiel	3
1.1	Différentiabilité	3
1.2	Gradient	5
1.3	Différentielle seconde et matrice hessienne	7
1.4	Formules de Taylor	9

Avant-propos

Ce polycopié est en partie tiré de ceux de nos prédécesseurs ou de collègues, cités ici pêle-mêle : Bruno Galerne, Joan Glaunès, Quentin Mérigot.

Voici quelques éléments de bibliographie.

- [R] François ROUVIÈRE, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*, troisième édition, Cassini, 2009
- [C] Philippe G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, cinquième édition, Dunod, 1998
- [BV] Stephen BOYD and Lieven VANDENBERGHE *Convex Optimization*, Cambridge University Press, 2004.
L'ouvrage peut être téléchargé gratuitement à l'adresse <http://stanford.edu/~boyd/cvxbook/>.
- [AK] Grégoire ALLAIRE et Sidi Mahmoud KABER, *Algèbre linéaire numérique*, Ellipses, 2002

Ce cours est une introduction aux problèmes d'optimisation. Le cours se focalise sur des problèmes d'optimisation sans ou avec contraintes pour les fonctions suffisamment différentiables, en dimension finie. Outre une introduction aux différentes notions mathématiques nécessaires (rappels de calcul différentiel, conditions d'optimalité, convexité, etc.), une part importante est donnée à l'exposition de différents algorithmes classiques d'optimisation, l'étude théorique de leur convergence, ainsi que leur mise en œuvre pratique. Les séances de TP se feront à l'aide du langage Python.

Cadre et notations

Espaces. L'ensemble du cours se déroulera dans le cadre d'espaces vectoriels normés de dimension finie. Pour $n \geq 1$, on notera $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$, ainsi que $\langle \cdot, \cdot \rangle$ le produit scalaire euclidien sur $\mathbb{R}^n$ et $\|\cdot\|$ la norme associée. En d'autres termes, pour tous $x, y \in \mathbb{R}^n$,

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i, \quad \|x\|^2 = \langle x, x \rangle = \sum_{i=1}^n x_i^2.$$

Pour $x \in \mathbb{R}^n$, la notation $B(x, r)$ (resp. $\overline{B}(x, r)$) fait référence à la boule ouverte (resp. fermée) de centre x et de rayon r pour la norme euclidienne.

Matrices. On note $\mathcal{M}_{m,n}(\mathbb{R})$ l'ensemble des matrices de taille $m \times n$ (m lignes, n colonnes) à coefficients réels, ainsi que $\mathcal{M}_n(\mathbb{R}) = \mathcal{M}_{n,n}(\mathbb{R})$ l'ensemble des matrices carrées de taille $n \times n$. On fera très souvent l'identification entre un vecteur de $\mathbb{R}^n$ et le vecteur colonne correspondant (c'est-à-dire la matrice de $\mathcal{M}_{n,1}(\mathbb{R})$).

La transposée d'une matrice $A \in \mathcal{M}_{m,n}(\mathbb{R})$ est notée $A^T \in \mathcal{M}_{n,m}(\mathbb{R})$.¹ Ainsi, avec les notations et identifications précédentes, on a pour tous $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^T y$ et par conséquent, pour tous

1. La notation A^T est standard dans le monde anglo-saxon et s'est imposée dans la littérature scientifique. Les étudiantes et étudiants sont libres d'utiliser la notation franco-française ${}^t A$ pour leurs prises de notes et leurs copies d'examen.

$A \in \mathcal{M}_{m,n}(\mathbb{R})$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

$$x^T A^T y = \langle Ax, y \rangle = \langle x, A^T y \rangle = y^T Ax.$$

On notera $\mathcal{S}_n(\mathbb{R})$ l'ensemble des matrices symétriques réelles ainsi que $\mathcal{S}_n^+(\mathbb{R})$ (resp. $\mathcal{S}_n^{++}(\mathbb{R})$) l'ensemble des matrices symétriques réelles positives (resp. définies positives). Une matrice de $\mathcal{S}_n(\mathbb{R})$ est diagonalisable (en base orthonormée) sur $\mathbb{R}$ et est alors dans $\mathcal{S}_n^+(\mathbb{R})$ (resp. $\mathcal{S}_n^{++}(\mathbb{R})$) si et seulement si toutes ses valeurs propres sont positives (resp. strictement positives).

On rappelle que l'ensemble des matrices symétriques réelles $\mathcal{S}_n(\mathbb{R})$ peut être muni d'une relation d'ordre (partielle) définie par

$$A \preceq B \iff \forall x \in \mathbb{R}^n, x^T Ax \leq x^T Bx.$$

Pour vérifier qu'il s'agit en effet une relation d'ordre, la seule difficulté concerne l'antisymétrie (si $A, B \in \mathcal{S}_n(\mathbb{R})$ satisfont $A \preceq B$ et $B \preceq A$, alors $A = B$) qui repose la diagonalisabilité des matrices symétriques réelles. Cette relation d'ordre sera utilisée dans la suite du cours pour énoncer de manière concise certains résultats de convergence des algorithmes étudiés.

Applications linéaires et matrices associées. Si E et F sont des espaces vectoriels normés de dimension finie, on désigne par $L(E, F)$ l'ensemble des applications linéaires (continues) de E dans F .

Dans le cas où $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ et vue l'identification entre vecteurs de $\mathbb{R}^n$ et vecteurs colonnes, on pourra donc écrire pour $\varphi \in L(\mathbb{R}^n, \mathbb{R}^m)$ l'égalité $\varphi(x) = Ax$, où $A \in \mathcal{M}_{m,n}(\mathbb{R})$ est la matrice de l'application φ dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$.

1 Rappels de calcul différentiel

1.1 Différentiabilité

Dans la suite de ce chapitre, E et F feront toujours référence à des espaces vectoriels de dimension finie, munis d'une norme quelconque, toutes les normes étant équivalentes dans ce cadre de dimension finie. On notera indifféremment $\|\cdot\|$ les normes impliquées (mais dans le cadre de $\mathbb{R}^n$, cette notation fera généralement référence à la norme euclidienne).

Enfin, Ω désignera systématiquement un ensemble *ouvert* de E .

1.1.1 Différentielle

Définition 1.1. Soient $f : \Omega \rightarrow F$ et $x \in \Omega$. On dit que f est *différentiable* au point $x \in \Omega$ s'il existe une application linéaire (continue) $\ell \in L(E, F)$ telle que, pour tout h dans un voisinage de 0, on ait

$$f(x + h) = f(x) + \ell(h) + o(h)$$

Dans ce cas ℓ est unique, est appelée *différentielle* de f au point x , et notée $df(x)$.

On dit que f est *différentiable* sur Ω si f est différentiable en tout point de Ω , et qu'elle est de classe C^1 si de plus $df : x \in \Omega \mapsto df(x)$ est continue sur Ω .

Par définition, la notation $o(h)$ fait référence au fait que l'on peut écrire

$$o(h) = \|h\| \varepsilon(h),$$

où ε est une fonction définie sur un voisinage de 0 et à valeurs dans F , qui tend vers 0 quand $h \rightarrow 0$. Cela est donc encore équivalent au fait que

$$\frac{o(h)}{\|h\|} \xrightarrow{h \rightarrow 0} 0.$$

La fonction $y \in E \mapsto f(x) + df(x)(y - x)$ représente la meilleure approximation affine de f au voisinage du point x .

Pratique de calcul de différentielle : première approche. Pour $x \in \Omega$ fixé, on écrit $f(x + h) - f(x)$ sous la forme de deux termes $\ell(h) + r(h)$, le premier terme $\ell(h)$ étant linéaire en h , et le second terme qui soit un $o(h)$. En général, la partie linéaire "saute aux yeux", seul le fait que le reste soit un $o(h)$ peut nécessiter un peu de travail, il s'agit de montrer que

$$\frac{r(h)}{\|h\|} \xrightarrow{h \rightarrow 0} 0,$$

ce qui se fait en général par le biais de majorations. On conclut alors que f est différentiable en x , et de plus que $df(x) = \ell$.

Exemple. Soient A une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$. Montrer que la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ définie par $f(x) = Ax + b$ est différentiable sur $\mathbb{R}^n$ et calculer sa différentielle en tout point $x \in \mathbb{R}^n$. En déduire que f est de classe C^1 sur $\mathbb{R}^n$.

Corrigé : Pour tous $x, h \in \mathbb{R}^n$,

$$f(x + h) = A(x + h) + b = f(x) + Ah + 0$$

L'application $h \mapsto Ah$ est linéaire, et le reste est nul (et son quotient par $\|h\|$ tend bien vers 0 quand $h \rightarrow 0$). Ainsi, donc f est différentiable en x et $df(x)$ est l'application $h \mapsto Ah$. Cela étant vrai pour tout $x \in \mathbb{R}^n$, f est différentiable sur $\mathbb{R}^n$. Puis comme l'application différentielle $df : x \in \mathbb{R}^n \mapsto df(x)$ est constante, c'est en particulier une application continue, donc f est bien de classe C^1 sur $\mathbb{R}^n$.

Exemple. Soit $f : I \rightarrow \mathbb{R}$ dérivable sur I , intervalle ouvert de $\mathbb{R}$. Montrer que f est différentiable sur I et exprimer la différentielle de f en tout point $x \in I$, en fonction de la dérivée f' de f . Quid du cas plus général où $f : I \rightarrow \mathbb{R}^m$?

Corrigé : Soit $x \in I$. Comme f est dérivable en x , par définition $f(x + h) = f(x) + f'(x)h + o(h)$. Comme $h \in \mathbb{R} \mapsto f'(x)h$ est linéaire (continue), f est bien différentiable en x et pour tout $h \in \mathbb{R}$, $df(x)(h) = f'(x)h$. Tout ce qui a été dit est en réalité valable pour $f : \mathbb{R} \rightarrow \mathbb{R}^m$ avec $m \in \mathbb{N}^*$ quelconque.

Par exemple, l'application $\cos$ est dérivable sur $\mathbb{R}$, elle est donc différentiable sur $\mathbb{R}$ et pour tout $x \in \mathbb{R}$, tout $h \in \mathbb{R}$, $d(\cos)(x)(h) = (-\sin(x))h = -\sin(x)h$.

Terminons par un dernier résultat à propos du lien entre différentiabilité et dérivées directionnelles.

Proposition 1.2 (*Dérivée directionnelle*)

On suppose que $f : \Omega \rightarrow F$ est différentiable en $x \in \Omega$. Alors pour tout $h \in E$,

$$\frac{f(x + th) - f(x)}{t} \xrightarrow[t \rightarrow 0]{} df(x)(h).$$

Cette égalité s'écrit encore

$$f(x + th) = f(x) + t df(x)(h) + o(t).$$

1.1.2 Jacobienne

Plaçons-nous désormais plus spécifiquement dans le cas important où $E = \mathbb{R}^n$, $F = \mathbb{R}^m$, si bien qu'on peut écrire $f = (f_1, \dots, f_m)$ avec $f_i : \Omega \rightarrow \mathbb{R}$.

Si f est différentiable en $x \in \Omega$ alors pour tout $j \in \{1, \dots, n\}$, la fonction f admet une j -ème dérivée partielle en x (que l'on notera dans ce cours $\partial_j f$), et

$$df(x)(e_j) = \partial_j f(x) = \lim_{t \rightarrow 0} \frac{f(x + te_j) - f(x)}{t}.$$

Définition 1.3. Si f est différentiable en $x \in \Omega$, la matrice de $df(x)$ (dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$) est appelée *matrice jacobienne* de f au point x , notée $J_f(x)$.

Celle-ci est donc donnée par

$$J_f(x) = (\partial_j f_i(x))_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{M}_{m,n}(\mathbb{R}).$$

Sous ces hypothèses, l'application de $df(x)$ à un vecteur $h \in \mathbb{R}^n$ se calcule donc via

$$df(x)(h) = \sum_{j=1}^n \partial_j f(x) h_j = J_f(x)h,$$

1.2 Gradient

Dans ce cours, notre obsession sera de minimiser des fonctions, qui seront donc à valeurs réelles ($m = 1$ dans ce qui précède).

Définition 1.4 (Gradient d'une fonction). Si $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est différentiable en $x \in \Omega$ alors $df(x) \in L(\mathbb{R}^n, \mathbb{R})$ est une forme linéaire sur $\mathbb{R}^n$; il existe donc un unique vecteur, noté $\nabla f(x) \in \mathbb{R}^n$, tel que

$$\forall h \in \mathbb{R}^n, \quad df(x)(h) = \langle \nabla f(x), h \rangle,$$

appelé *gradient* de f en x .

De manière équivalente, la matrice $J_f(x)$ est une matrice ligne (de taille $1 \times n$) et la transposée de cette matrice est $\nabla f(x)$. Pour tout $h \in \mathbb{R}^n$,

$$df(x)(h) = \nabla f(x)^T h = \langle \nabla f(x), h \rangle.$$

Enfin, $\nabla f(x)$ se calcule à partir des dérivées partielles² via

$$\nabla f(x) = (\partial_i f(x))_{1 \leq i \leq n} \in \mathbb{R}^n.$$

Ainsi, pour $f : \Omega \rightarrow \mathbb{R}$, f est différentiable en $x \in \Omega$ si et seulement si on peut écrire

$$f(x+h) = f(x) + \langle v, h \rangle + o(h),$$

pour un certain vecteur v , auquel cas $v = \nabla f(x)$.

Interprétation du vecteur gradient. $\nabla f(x)$ s'interprète comme la direction de plus forte augmentation de f en x ; sa norme donne le taux de variation de f dans cette direction. Par ailleurs, $\nabla f(x)$ est orthogonal aux lignes de niveau de la fonction f .

Pratique de calcul de différentielle/jacobienne/gradient : seconde approche. D'après vos cours de Licence, les fonctions usuelles (polynomiales, exponentielle, etc) sont de classe C^∞ ³ sur leurs domaines de définition. Ainsi, on sait déjà a priori que ces fonctions sont différentiables. La jacobienne peut alors se déterminer par simple calcul des dérivées partielles, ce qui est la méthode recommandée si la fonction est une "brave" fonction explicite, comme par exemple, $f : (x, y) \mapsto e^{x^2+3yx+2y^4}$. Dans ce cas, on s'évitera la souffrance inutile d'une décomposition de $f(x+h) - f(x)$.

Exercice 1

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$\forall x \in \mathbb{R}^n, \quad f(x) = \frac{1}{2} \langle Ax, x \rangle,$$

où A est une matrice de $\mathcal{M}_n(\mathbb{R})$.

1. Montrer que f est différentiable sur $\mathbb{R}^n$ et déterminer $\nabla f(x)$ pour tout x .
2. Quel est le gradient de l'application $x \mapsto \frac{1}{2} \|x\|^2$?

Corrigé :

1. Pour tous $x, h \in \mathbb{R}^n$,

$$\begin{aligned} f(x+h) &= \frac{1}{2} \langle A(x+h), x+h \rangle \\ &= \frac{1}{2} \langle Ax, x \rangle + \frac{1}{2} \langle Ax, h \rangle + \frac{1}{2} \langle Ah, x \rangle + \frac{1}{2} \langle Ah, h \rangle \\ &= f(x) + \frac{1}{2} \langle (A + A^T)x, h \rangle + \frac{1}{2} \langle Ah, h \rangle. \end{aligned}$$

Or, par l'inégalité de Cauchy-Schwarz,

$$\left| \frac{1}{2} \langle Ah, h \rangle \right| \leq \frac{1}{2} \|Ah\| \|h\| \leq \frac{1}{2} \|A\| \|h\|^2,$$

donc ce terme peut bien s'écrire $o(h)$. Ainsi, f est différentiable en tout point $x \in \mathbb{R}^n$ et

$$\nabla f(x) = \frac{1}{2} (A + A^T)x.$$

2. En pratique, dans ce cours, nous calculerons que très rarement le gradient via les dérivées partielles.
3. Voir plus bas pour des rappels sur la différentiabilité d'ordre supérieur.

Ici, on aurait pu affirmer directement que f est différentiable en tant que fonction polynomiale (quadratique). Mais le calcul du gradient par dérivées partielles s'avère plus pénible. Pour calculer $\partial_k f$ pour un certain $k \in \{1, \dots, n\}$, on peut écrire pour $x \in \mathbb{R}^n$

$$f(x) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \frac{1}{2} \left(a_{kk} x_k^2 + x_k \sum_{j \neq k} a_{kj} x_j + x_k \sum_{i \neq k} a_{ik} x_i \right) + g(x),$$

où g est une fonction qui ne dépend pas de x_k . Un calcul direct de dérivée donne alors

$$\partial_k f(x) = a_{kk} x_k + \frac{1}{2} \left(\sum_{j \neq k} a_{kj} x_j + \sum_{i \neq k} a_{ik} x_i \right) = \frac{1}{2} \left(\sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n a_{ik} x_i \right),$$

cette dernière expression étant bien la k -ème composante du vecteur $\frac{1}{2}(A + A^T)x$, c'est-à-dire qu'on retrouve l'identité

$$\nabla f(x) = (\partial_1 f(x), \dots, \partial_n f(x))^T = \frac{1}{2}(A + A^T)x.$$

2. Il s'agit du cas particulier où $A = \text{Id}$, pour lequel on trouve $\nabla f(x) = x$.

1.2.1 Différentiation de fonctions composées

Théorème 1.5 (*Différentiation de fonctions composées*)

Soient $f : \Omega \rightarrow \mathbb{R}^m$ et $g : \Omega' \rightarrow \mathbb{R}^p$ deux fonctions définies sur des ouverts $\Omega \subset \mathbb{R}^n$ et $\Omega' \subset \mathbb{R}^m$, avec $f(\Omega) \subset \Omega'$. Si f est différentiable en $x \in \Omega$ et g est différentiable en $f(x) \in \Omega'$, alors $h = g \circ f : \Omega \rightarrow \mathbb{R}^p$ est différentiable en x et de plus

$$dh(x) = dg(f(x)) \circ df(x).$$

En terme de matrices jacobiniennes, cela s'écrit

$$J_h(x) = J_g(f(x))J_f(x).$$

Exemple. Déterminons le gradient de l'application $h : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$\forall x \in \mathbb{R}^n, \quad h(x) = g(Ax + b)$$

où A est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ et $g : \mathbb{R}^m \rightarrow \mathbb{R}$ est une application différentiable sur $\mathbb{R}^m$. On a $h(x) = g \circ f(x)$ avec $f(x) = Ax + b$. Comme f est affine, $J_f(x) = A$ en tout $x \in \mathbb{R}^n$. On a donc d'après la règle de dérivation des fonctions composées

$$J_h(x) = J_g(f(x))J_f(x) = J_g(Ax + b)A.$$

Donc $\nabla h(x) = J_h(x)^T = A^T J_g(Ax + b)^T = A^T \nabla g(Ax + b)$.

1.3 Différentielle seconde et matrice hessienne

On s'intéresse aux différentiabilités d'ordre supérieur, dans le cas d'applications $f : \Omega \rightarrow \mathbb{R}$, où Ω est un ouvert de $\mathbb{R}^n$. Si f est différentiable sur Ω , on peut s'intéresser à la différentiabilité de l'application $df : \Omega \rightarrow L(\mathbb{R}^n, \mathbb{R})$, le rôle d'espace vectoriel (de dimension finie égale à n) d'arrivée étant donc ici joué par $L(\mathbb{R}^n, \mathbb{R})$.

Définition 1.6 (Différentielle seconde). Soit $f : \Omega \rightarrow \mathbb{R}$ une application différentiable sur Ω . On dit que f est deux fois différentiable en $a \in \Omega$ si l'application

$$df : x \mapsto df(x),$$

est différentiable en x . On note cette différentielle $d^2f(x) = d(df)(x)$ et on l'appelle la différentielle seconde de f en x .

On dit que f est de classe C^2 sur Ω si f est deux fois différentiable en tout point de Ω et si de plus $x \mapsto d^2f(x)$ est continue sur Ω .

Comme $df : \Omega \subset \mathbb{R}^n \rightarrow L(\mathbb{R}^n, \mathbb{R})$, on a ainsi $d^2f(x) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}))$. Par conséquent, si $x \in \Omega$ et $h \in \mathbb{R}^n$, $df(x)(h)$ est un élément de $L(\mathbb{R}^n, \mathbb{R})$, que l'on peut évaluer en $k \in \mathbb{R}^n$. L'application

$$(h, k) \mapsto d^2f(x)(h)(k),$$

est alors une application bilinéaire (continue) de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$, à laquelle on identifie $d^2f(x)$; cette identification permet d'écrire $d^2f(x)(h, k)$, notation qui a le mérite d'être plus compacte.

Théorème 1.7 (Théorème de Schwarz)

Soit $f : \Omega \rightarrow \mathbb{R}$ une application deux fois différentiable en $x \in \Omega$. Alors $d^2f(x) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ est une application bilinéaire symétrique, c'est-à-dire

$$\forall (h, k) \in \mathbb{R}^n \times \mathbb{R}^n, \quad d^2f(x)(h, k) = d^2f(x)(k, h).$$

La différentielle seconde en un point est alors une *forme* bilinéaire symétrique que l'on peut caractériser par sa matrice dans la base canonique de $\mathbb{R}^n$.

Définition 1.8 (Matrice hessienne). Soit $f : \Omega \rightarrow \mathbb{R}$ une application deux fois différentiable en $x \in \Omega$. On peut alors considérer la forme bilinéaire symétrique $d^2f(x)$, et sa matrice dans la base canonique de $\mathbb{R}^n$, appelée *matrice hessienne de f en x* et notée

$$\nabla^2 f(x) = (d^2f(x)(e_i, e_j))_{1 \leq i, j \leq n} = (\partial_{ij}^2 f(x))_{1 \leq i, j \leq n}.$$

Par définition, on dispose de l'écriture

$$\forall (h, k) \in \mathbb{R}^n \times \mathbb{R}^n, \quad d^2f(x)(h, k) = \langle \nabla^2 f(x)h, k \rangle = k^T \nabla^2 f(x)h = h^T \nabla^2 f(x)k.$$

Dans ce cours, notre pain quotidien sera de calculer le gradient et la matrice hessienne de fonctions $f : \Omega \rightarrow \mathbb{R}$ deux fois différentiables. En pratique, on utilise la proposition suivante selon laquelle la matrice hessienne est la jacobienne du gradient.

Proposition 1.9 (La matrice hessienne est la jacobienne du gradient)

Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction deux fois différentiable au point $x \in \Omega$. Alors, la matrice hessienne $\nabla^2 f(x)$ de f au point x est la matrice jacobienne $J_{\nabla f}(x)$ de l'application gradient $x \mapsto \nabla f(x)$ au point x . En d'autres termes,

$$\forall h \in \mathbb{R}^n, \quad d(\nabla f)(x)(h) = J_{\nabla f}(x)h = \nabla^2 f(x)h.$$

Notons que la première égalité ci-dessus est simplement la définition de la Jacobienne, c'est la deuxième qui constitue véritablement la substance de cette proposition.

Démonstration : Posons $g(x) = \nabla f(x)$ pour tout $x \in \Omega$. Alors, par définition, la différentielle de $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est caractérisée par sa matrice jacobienne

$$J_g(x) = (\partial_j g_i(x))_{1 \leq i, j \leq n} = (\partial_j (\partial_i f)(x))_{1 \leq i, j \leq n} = (\partial_{ij}^2 f(x))_{1 \leq i, j \leq n} = \nabla^2 f(x). \quad \blacksquare$$

Exemple. Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$h(x) = g(Ax + b)$$

où A est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ et $g : \mathbb{R}^m \rightarrow \mathbb{R}$ est une application deux fois différentiable sur $\mathbb{R}^m$. Alors, par composée, h est deux fois différentiable et

$$\nabla^2 h(x) = A^T \nabla^2 g(Ax + b) A,$$

Cette formule s'obtient facilement en utilisant la proposition précédente puis en prenant la jacobienne de $x \in \mathbb{R}^n \mapsto \nabla h(x)$ dont l'expression $\nabla h(x) = A^T \nabla g(Ax + b)$ a été montrée précédemment.

1.4 Formules de Taylor

Les formules de Taylor se généralisent aux fonctions de plusieurs variables. On se limite aux fonctions à valeurs réelles et on donne les formules dans un cadre de régularité non optimale : les formules de Taylor-Young d'ordre 1 (resp. d'ordre 2) sont bien entendu valables dès que f est une fois (resp. deux fois) différentiable en le point d'intérêt.

Théorème 1.10 (Formules de Taylor pour les fonctions de classe C^1)

Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 sur Ω .

1. La définition de la différentielle donne la formule de Taylor-Young à l'ordre 1 :

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \|h\| \varepsilon(h) \quad \text{avec} \quad \lim_{h \rightarrow 0} \varepsilon(h) = 0.$$

On considère maintenant un point h fixé tel que le segment $[x, x + h]$ soit inclus dans Ω .

2. Inégalité des accroissements finis :

$$|f(x + h) - f(x)| \leq \sup_{y \in [x, x+h]} \|\nabla f(y)\| \|h\|.$$

3. Formule de Taylor-Maclaurin : il existe $\theta \in]0, 1[$ tel que

$$f(x + h) = f(x) + \langle \nabla f(x + \theta h), h \rangle.$$

4. Formule de Taylor avec reste intégral :

$$f(x + h) = f(x) + \int_0^1 \langle \nabla f(x + th), h \rangle dt.$$

Notons que ces formules de Taylor multidimensionnelles s'obtiennent par application des formules de Taylor usuelles (i.e. pour les fonctions de la variable réelle) à la fonction $t \mapsto f(x + th)$ sur $[0, 1]$.

Théorème 1.11 (*Formules de Taylor pour les fonctions de classe C^2*)

Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^2 sur Ω .

1. Formule de Taylor-Young à l'ordre 2 :

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \|h\|^2 \varepsilon(h) \quad \text{avec } \lim_{h \rightarrow 0} \varepsilon(h) = 0.$$

On considère maintenant un point h fixé tel que le segment $[x, x+h]$ soit inclus dans Ω .

2. Formule des accroissements finis généralisée :

$$|f(x+h) - f(x) - \langle \nabla f(x), h \rangle| \leq \frac{1}{2} \sup_{y \in]x, x+h[} \|\nabla^2 f(y)\|_{\mathcal{M}_n(\mathbb{R})} \|h\|^2.$$

où $\|\cdot\|_{\mathcal{M}_n(\mathbb{R})}$ désigne la norme subordonnée à la norme euclidienne.

3. Formule de Taylor-Maclaurin : il existe $\theta \in]0, 1[$ tel que

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x+\theta h) h, h \rangle.$$

4. Formule de Taylor avec reste intégral :

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \int_0^1 (1-t) \langle \nabla^2 f(x+th) h, h \rangle dt.$$

Une référence pour ce chapitre est le chapitre 8 de [C].

Références

- [R] François ROUVIÈRE, *Petit guide du calcul différentiel à l'usage de la licence et de l'agrégation*, Cassini, 2009
- [AK] Grégoire ALLAIRE et Sidi Mahmoud KABER, *Algèbre linéaire numérique*, Ellipses, 2002
- [C] Philippe G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, cinquième édition, Dunod, 1998
- [BV] Stephen BOYD and Lieven VANDENBERGHE *Convex Optimization*, Cambridge University Press, 2004