Université Paris Dauphine
DUMI2E 2e année
Année 2008-2009

Calcul différentiel et optimisation I

François BOLLEY

Chapitre 1 - L'espace $\mathbb{R}^n$

Soit $\mathbb{R}$ l'ensemble des nombres réels. Cet ensemble $\mathbb{R}$ est muni de structures algébrique et topologique construites à partir des opérations d'addition $x+y$ et de multiplication xy, de la valeur absolue $|x|$ et de la relation d'ordre $x \leq y$ avec les propriétés et notations classiques que l'on ne rappellera pas. Dans le cadre de ces structures on a étudié certains aspects des fonctions d'une variable réelle à valeurs réelles

$$f : U \subset \mathbb{R} \to \mathbb{R} : x \mapsto f(x).$$

Pour $n \geq 1$, soit $\mathbb{R}^n$ l'ensemble des n-uplets de nombres réels de la forme

$$x = (x_1, \ldots, x_n)$$

où $x_1, \ldots, x_n$ sont des nombres réels appelés les composantes de x.

On se propose d'étudier dans ce cours certains aspects des fonctions de n variables réelles à valeurs vectorielles

$$f : U \subset \mathbb{R}^n \to \mathbb{R}^p : x \mapsto f(x)$$

et pour cela dans ce premier chapitre, on va définir sur ces ensembles $\mathbb{R}^n$ et $\mathbb{R}^p$ des structures qui prolongent les structures de l'ensemble $\mathbb{R}$.

1. Propriétés algébriques de $\mathbb{R}^n$

1.1. Structure d'espace vectoriel sur $\mathbb{R}^n$

Pour x et $y \in \mathbb{R}^n$ on définit

$$x + y = (x_1 + y_1, \ldots, x_n + y_n)$$

que l'on appelle l'addition de x et y, et pour $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$ on définit

$$\lambda\, x = (\lambda x_1, \ldots, \lambda x_n)$$

que l'on appelle la multiplication de x par le scalaire λ.

Pour $i = 1, \ldots, n$ on définit l'élément e_i de $\mathbb{R}^n$ de composantes $e_{ij} = 0$ si $j \neq i$ et $e_{ii} = 1$.

Proposition 1. *Muni des deux opérations d'addition et de multiplication par un scalaire, l'ensemble $\mathbb{R}^n$ est un espace vectoriel sur le corps $\mathbb{R}$ de dimension n dont une base, dite base canonique, est formée par la famille $\{e_1, \ldots, e_n\}$.*

L'élément neutre pour l'addition $(0, \ldots, 0)$ sera noté simplement 0 et un élément de $\mathbb{R}^n$ sera appelé soit un vecteur, soit un point.

Tout élément x de $\mathbb{R}^n$ se décompose donc de manière unique sous la forme

$$x = x_1 e_1 + \ldots + x_n e_n$$

pour des cœfficients réels $x_1, \ldots, x_n$: ces cœfficients x_i sont en fait les composantes de x qu'on appellera aussi les composantes de x sur la base canonique de $\mathbb{R}^n$.

Ainsi un élément x de $\mathbb{R}^n$ sera représenté suivant le contexte
- soit par ses composantes sous la forme du n-uplet

$$x = (x_1, \ldots, x_n)$$

- soit par sa décomposition sur la base canonique de $\mathbb{R}^n$ sous la forme de la combinaison linéaire

$$x = x_1 e_1 + \ldots + x_n e_n.$$

1.2. Applications linéaires

Avec la structure algébrique de $\mathbb{R}^n$ définie précédemment, les applications linéaires jouent un rôle particulier.

Définition 2. *Une application f de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est dite linéaire si pour $x,\ y \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$ on a*

$$f(x+y) = f(x) + f(y)$$
$$f(\lambda\, x) = \lambda\, f(x).$$

Une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$ est dite une forme linéaire.

L'ensemble des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est noté $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$.

La linéarité est conservée par les opérations de multiplication par un scalaire, d'addition, de composition et d'inversion des applications linéaires. Pour l'inversibilité on a le résultat particulier suivant :

Proposition 3. *Une application linéaire f de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est bijective si et seulement si elle est injective, et si et seulement si elle est surjective.*

Dans ce cas l'application inverse, dite aussi réciproque, f^{-1} est alors une application linéaire de $\mathbb{R}^n$ sur $\mathbb{R}^n$, et on dit que f est inversible ou est un isomorphisme de $\mathbb{R}^n$ sur $\mathbb{R}^n$.

L'application inverse f^{-1} est l'unique application g de $\mathbb{R}^n$ dans $\mathbb{R}^n$ telle que

$$f \circ g = g \circ f = Id$$

où Id est l'application identité de $\mathbb{R}^n$ dans $\mathbb{R}^n$.

1.3. Matrices

Une matrice A de taille (p, n) est un tableau à p lignes et n colonnes de cœfficients réels de la forme

$$A = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & \vdots & \vdots \\ A_{p1} & \ldots & A_{pn} \end{bmatrix}$$

qu'on note aussi

$$A = \big[A_{ij}\big]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}} \quad \text{ou} \quad A = \big[A_{ij}\big]$$

le premier indice i étant celui de la ligne et le second j celui de la colonne.

On définit les opérations de multiplication par un scalaire, d'addition, de multiplication et d'inversion des matrices.

En particulier le produit d'une matrice A de taille (p,n) et d'une matrice B de taille (n,q) est la matrice notée AB de taille (p,q) dont les cœfficients sont définis par

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

Exemple. Etant données une matrice colonne X de taille $(n,1)$ de la forme

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

et une matrice A de taille (p,n), la matrice produit AX est la matrice colonne de taille $(p,1)$ donnée par

$$AX = \begin{bmatrix} A_{11}x_1 + ... + A_{1n}x_n \\ \vdots \\ A_{p1}x_1 + ... + A_{pn}x_n \end{bmatrix}.$$

Quant à l'inversibilité des matrices, en notant I la matrice unité de taille (n,n) telle que $I_{ij} = 0$ pour $i \neq j$ et $I_{ii} = 1$, on pose la définition suivante :

Définition 4. *Une matrice A de taille (n,n) est dite inversible s'il existe une matrice B de taille (n,n) telle que*

$$AB = BA = I.$$

La matrice B est alors unique, notée A^{-1} et appelée inverse de A.

Cette propriété peut se traduire à l'aide du déterminant de la matrice A. Le déterminant d'une matrice A de taille (n,n) est le nombre réel noté $\text{dét}\, A$ et défini par

$$\text{dét}\, A = \sum_{\sigma \in \Sigma_n} \varepsilon_\sigma \, A_{\sigma(1)1} \dots A_{\sigma(n)n}$$

où Σ_n est le groupe des permutations de l'ensemble $\{1, \dots, n\}$ et ε_σ est la signature de la permutation σ.

Proposition 5. *Une matrice A de taille (n,n) est inversible si et seulement si $\text{dét}\, A \neq 0$.*

Dans ce cas, la matrice inverse est donnée par

$$A^{-1} = \left(\text{dét}\, A\right)^{-1} \tilde{A}$$

où $\tilde{A}$ est la matrice de taille (n,n) de cœfficients $\tilde{A}_{ij} = (-1)^{i+j}\, \text{dét}\, a_{ji}$ et où a_{ji} est la matrice de taille $(n-1, n-1)$ obtenue en supprimant la ligne j et la colonne i de la matrice A .

1.4. Représentations matricielles

On convient d'identifier le vecteur $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$ et la matrice colonne de taille $(n, 1)$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

et de dire que cette matrice colonne est la représentation matricielle du vecteur $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$ relativement à la base canonique de $\mathbb{R}^n$.

Suivant le contexte, on pourra noter par la même lettre x soit le n-uplet $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$, soit la matrice colonne de taille $(n, 1)$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

et on pourra utiliser la représentation matricielle du vecteur à la place du vecteur lui-même.

On note plus précisément $\{e_j^n;\ j = 1, \ldots, n\}$ la base canonique de $\mathbb{R}^n$ et $\{e_i^p;\ i = 1, \ldots, p\}$ la base canonique de $\mathbb{R}^p$.

Si y_j pour $j = 1, ..., n$ sont n vecteurs de $\mathbb{R}^p$ de composantes A_{ij} pour $i = 1, ..., p$ dans la base $\{e_i^p;\ i = 1, \ldots, p\}$ de $\mathbb{R}^p$, c'est-à-dire de la forme

$$y_j = \sum_{i=1}^{p} A_{ij} e_i^p$$

on note A la matrice $A = \big[A_{ij}\big]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}}$ appelée la matrice représentative de la famille $\{y_1, ..., y_n\}$ dans la base $\{e_i^p;\ i = 1, \ldots, p\}$ de $\mathbb{R}^p$.

Si x est un vecteur de $\mathbb{R}^n$ et f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, on a par la linéarité de f

$$f(x) = \sum_{j=1}^{n} x_j\, f(e_j^n) = \sum_{i=1}^{p} \Big(\sum_{j=1}^{n} A_{ij} x_j \Big) e_i^p$$

où pour $j = 1, \ldots, n$, on a noté pour $i = 1, \ldots, p$

$$A_{ij} = \big(f(e_j^n)\big)_i$$

les p composantes du vecteur $f(e_j^n)$ de $\mathbb{R}^p$ sur la base canonique $\{e_i^p;\ i = 1, \ldots, p\}$ de $\mathbb{R}^p$.

Autrement dit la matrice $A = \big[A_{ij}\big]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}}$ est la matrice représentative de la famille $\{f(e_1^n), \cdots, f(e_n^n)\}$ dans la base $\{e_i^p;\ i = 1, \ldots, p\}$ de $\mathbb{R}^p$.

Par conséquent l'application f est déterminée de façon unique par la matrice A de taille (p, n) dont les cœfficients sont ces A_{ij}.

On peut donc énoncer :

Proposition 6. *Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. Alors il existe une unique matrice A de taille (p,n) telle que pour $x \in \mathbb{R}^n$ la matrice produit Ax est la représentation matricielle du vecteur $f(x)$ de $\mathbb{R}^p$ relativement à la base canonique de $\mathbb{R}^p$.*

Cette matrice A est appelée la matrice représentative de l'application linéaire f relativement aux bases canoniques de $\mathbb{R}^n$ et de $\mathbb{R}^p$.

Autrement dit on pourra noter

$$f(x) = Ax = \begin{bmatrix} A_{11}x_1 + ... + A_{1n}x_n \\ \vdots \\ A_{p1}x_1 + ... + A_{pn}x_n \end{bmatrix}.$$

Notant de façon générale $A(f)$ la matrice représentative de l'application linéaire f, on a dans des cadres convenables

1 - $A(\lambda f) = \lambda\, A(f)$,

2 - $A(f+g) = A(f) + A(g)$,

3 - $A(f \circ g) = A(f)A(g)$,

4 - f est un isomorphisme de $\mathbb{R}^n$ sur $\mathbb{R}^n$ si et seulement si $A(f)$ est inversible, et dans ce cas $A(f^{-1}) = \big(A(f)\big)^{-1}$.

1.5. Formes bilinéaires - Formes quadratiques

Définition 7. *Une forme bilinéaire f sur $\mathbb{R}^n \times \mathbb{R}^n$ est une application*

$$f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} : (x,y) \mapsto f(x,y)$$

telle que pour tous a et $b \in \mathbb{R}^n$, les applications partielles

$$f(a,\cdot) : \mathbb{R}^n \to \mathbb{R} : y \mapsto f(a,y)$$

$$f(\cdot,b) : \mathbb{R}^n \to \mathbb{R} : x \mapsto f(x,b)$$

soient linéaires.

La forme bilinéaire f est dite symétrique si $f(x,y) = f(y,x)$ pour x et $y \in \mathbb{R}^n$.

Le premier exemple fondamental de forme bilinéaire est le produit scalaire euclidien.

Définition 8. *Le produit scalaire euclidien des vecteurs x et y de $\mathbb{R}^n$ est le nombre réel noté $< x,y >$ et défini par*

$$< x,y >= x_1y_1 + \cdots + x_ny_n.$$

Le produit scalaire pourra aussi être écrit sous la forme d'un produit matriciel

$$< x,y > = x^T\, y = y^T\, x$$

où x^T est la matrice ligne de taille $(1,n)$

$$x^T = \big[x_1 \ldots x_n\big]$$

appelée matrice transposée de la matrice colonne de taille $(n, 1)$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

De façon générale la matrice transposée de la matrice A de taille (p, n) est la matrice de taille (n, p) notée A^T dont les cœfficients sont définis par

$$A^T_{ij} = A_{ji}.$$

Pour $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^p$ on a alors

$$< Ax, y >_p = < x, A^T y >_n$$

où $< \cdot, \cdot >_p$ est le produit scalaire euclidien de $\mathbb{R}^p$ et $< \cdot, \cdot >_n$ est le produit scalaire euclidien de $\mathbb{R}^n$.

Comme pour les applications linéaires on a la représentation matricielle suivante :

Proposition 9. *Soit f une forme bilinéaire sur $\mathbb{R}^n \times \mathbb{R}^n$. Alors il existe une unique matrice A de taille (n, n) telle que pour x et $y \in \mathbb{R}^n$ on ait*

$$f(x, y) = < Ax, y > .$$

Cette matrice A est alors appelée la matrice représentative de l'application bilinéaire f relativement à la base canonique de $\mathbb{R}^n$.

La forme bilinéaire f est symétrique si et seulement si sa matrice représentative A est symétrique, c'est-à-dire si $A_{ij} = A_{ji}$ pour $i, j = 1, \ldots, n$.

La matrice A est la matrice de taille (n, n) ayant pour cœfficients

$$A_{ij} = f(e_j, e_i).$$

Ainsi pour x et $y \in \mathbb{R}^n$

$$f(x, y) = \sum_{i,j=1}^{n} A_{ij} x_j y_i.$$

Exemple. La matrice représentative du produit scalaire euclidien est la matrice identité.

Définition 10. *Une forme quadratique f sur $\mathbb{R}^n$ est une application*

$$f : \mathbb{R}^n \to \mathbb{R} : x \mapsto f(x)$$

telle que

1 - $f(\lambda x) = \lambda^2 f(x)$ pour $\lambda \in \mathbb{R}$ et $x \in \mathbb{R}^n$,

2 - l'application

$$\phi : (x, y) \mapsto f(x + y) - f(x) - f(y)$$

est une forme bilinéaire sur $\mathbb{R}^n \times \mathbb{R}^n$.

On a la représentation matricielle suivante :

Proposition 11. *Soit f une forme quadratique sur $\mathbb{R}^n$. Alors il existe une unique matrice symétrique A de taille (n,n) telle que pour x et $y \in \mathbb{R}^n$ on ait*

$$f(x) = < Ax, x > .$$

Cette matrice A est alors appelée la matrice représentative de la forme quadratique f relativement à la base canonique de $\mathbb{R}^n$.

2. Propriétés topologiques de $\mathbb{R}^n$

2.1. Normes sur $\mathbb{R}^n$

Définition 12. *Une norme sur $\mathbb{R}^n$ est une application de $\mathbb{R}^n$ dans $\mathbb{R}$ notée*

$$\| \cdot \| : \mathbb{R}^n \to \mathbb{R} : x \mapsto \|x\|$$

telle que
1 - (positivité) $\|x\| \geq 0$ pour $x \in \mathbb{R}^n$ et $\|x\| = 0$ si et seulement si $x = 0$,
2 - (inégalité triangulaire) $\|x+y\| \leq \|x\| + \|y\|$ pour x et $y \in \mathbb{R}^n$,
3 - (homogénéité) $\|\lambda\, x\| = |\lambda|\, \|x\|$ pour $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$.

L'espace vectoriel $\mathbb{R}^n$ muni d'une norme $\| \cdot \|$ sera dit un espace vectoriel normé (e.v.n. en abrégé) et noté $(\mathbb{R}^n, \| \cdot \|)$.

En particulier l'inégalité triangulaire implique l'inégalité suivante souvent utilisée

$$|\, \|x\| - \|y\|\, | \leq \|x - y\|.$$

Définition 13. *Deux normes $\| \cdot \|$ et $\| \cdot \|^*$ sur $\mathbb{R}^n$ sont dites équivalentes s'il existe deux constantes C_1 et $C_2 > 0$ telles que pour $x \in \mathbb{R}^n$*

$$C_1\, \|x\| \leq \|x\|^* \leq C_2\, \|x\|.$$

Exemples. Pour $x = (x_1, \ldots, x_n)$ on pose

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|\,, \qquad \|x\|_2 = \Big(\sum_{i=1}^{n} x_i^2 \Big)^{\frac{1}{2}} = < x, x >^{\frac{1}{2}}\,, \qquad \|x\|_\infty = \sup_{i=1,\ldots,n} |x_i|.$$

Alors les applications $\| \cdot \|_p$ sont des normes sur $\mathbb{R}^n$ pour $p = 1,\ 2,\ \infty$ et ces normes sont équivalentes car pour tout $x \in \mathbb{R}^n$ on a

$$\frac{1}{\sqrt{n}} \|x\|_1 \leq \|x\|_2 \leq \|x\|_1,$$
$$\|x\|_\infty \leq \|x\|_1 \leq n\, \|x\|_\infty,$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\, \|x\|_\infty\,.$$

Pour la norme $\|\cdot\|_2$, appelée norme euclidienne, on a en particulier

Proposition 14. *Pour x et $y \in \mathbb{R}^n$ on a*
1 - (inégalité de Cauchy-Schwarz) $| < x, y > | \leq \|x\|_2 \, \|y\|_2$,
2 - (égalité) $| < x, y > | = \|x\|_2 \, \|y\|_2$ si et seulement si, pour $x \neq 0$, il existe $\lambda \in \mathbb{R}$ tel que $y = \lambda x$.

Démonstration. On suppose que $x \neq 0$ et donc $\|x\|_2 \neq 0$. Pour $\lambda \in \mathbb{R}$ on a

$$\|\lambda x + y\|_2^2 = \lambda^2 \|x\|_2^2 + 2\lambda < x, y > + \|y\|_2^2.$$

Par suite pour tout $\lambda \in \mathbb{R}$ on a

$$\lambda^2 \|x\|_2^2 + 2\lambda < x, y > + \|y\|_2^2 \geq 0.$$

Le discriminant de ce trinôme est donc nécessairement ≤ 0, ce qui implique l'inégalité de Cauchy-Schwarz.

Si l'égalité a lieu dans cette inégalité, c'est-à-dire si le discriminant du trinôme est nul, il existe alors $\lambda \in \mathbb{R}$ pour lequel ce trinôme est nul, c'est-à-dire $\|\lambda x + y\|_2^2 = 0$ soit $y = -\lambda x$.$\diamond$

Il est très commode de pouvoir traduire certaines propriétés d'un e.v.n. $(\mathbb{R}^n, \|\cdot\|)$ à l'aide de suites de points de $\mathbb{R}^n$. On suppose connues les propriétés sur les suites de nombres réels.

Définition 15. *Une suite de $\mathbb{R}^n$ est une application de l'ensemble $\mathbb{N}$ des nombres entiers naturels dans $\mathbb{R}^n$*

$$\mathbb{N} \to \mathbb{R}^n : k \mapsto x_k.$$

Une telle suite est notée $(x_k)_{k\geq 0}$ ou $(x_k)_k$ ou (x_k).

Si $(k_j)_{j\geq 0}$ est une suite strictement croissante de nombres entiers, la suite $(y_j)_{j\geq 0}$ définie par $y_j = x_{k_j}$ est dite une sous-suite ou suite extraite de la suite $(x_k)_{k\geq 0}$ et notée $(x_{k_j})_{j\geq 0}$ ou $(x_{k_j})_j$ ou (x_{k_j}).

Définition 16. *Soit $(x_k)_k$ une suite de $\mathbb{R}^n$. On dit que la suite $(x_k)_k$ est convergente dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$ s'il existe x appartenant à $\mathbb{R}^n$ tel que pour tout $\varepsilon > 0$ il existe un entier $K > 0$ tel que $\|x_k - x\| \leq \varepsilon$ pour tout $k \geq K$.*

Cet élément x est alors unique. On dit que la suite $(x_k)_k$ converge vers x dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$, que x est la limite de la suite $(x_k)_k$ et on note

$$\lim_{k\to\infty} x_k = x.$$

Autrement dit la suite $(\|x_k - x\|)_k$ converge vers 0 dans $(\mathbb{R}, |\cdot|)$.

On va étendre le théorème de Bolzano-Weierstrass pour les suites bornées de nombres réels au cas des suites de points de $\mathbb{R}^n$ bornées dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|_\infty)$.

Définition 17. *Une suite* $(x_k)_k$ *de points de* $\mathbb{R}^n$ *est dite bornée dans l'e.v.n.* $(\mathbb{R}^n, \|\cdot\|)$ *s'il existe une constante* $r > 0$ *telle que* $\|x_k\| \leq r$ *pour tout* $k \geq 0$.

Théorème 18 (de Bolzano-Weierstrass). *Toute suite* $(x_k)_k$ *de points de* $\mathbb{R}^n$ *bornée dans l'e.v.n.* $(\mathbb{R}^n, \|\cdot\|_\infty)$ *contient une sous-suite convergente* $(x_{k_j})_j$ *dans l'e.v.n.* $(\mathbb{R}^n, \|\cdot\|_\infty)$.

Démonstration. Soit $(x_k)_k$ une suite de points de $\mathbb{R}^n$ que l'on suppose bornée dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|_\infty)$, c'est-à-dire que pour une constante $r > 0$ on a

$$|x_{ki}| \leq r \quad \text{pour} \quad 1 \leq i \leq n \ \text{ et } k \geq 0$$

où on a noté $x_k = (x_{k1}, \ldots, x_{kn})$.

1 - Le cas $n = 1$ est classique. On rappelle une démonstration dans ce cas.

En notant $P = \{k \in \mathbb{N};\ x_{k+i} \leq x_k,\ i = 0, 1, \ldots\}$, deux cas sont à examiner.
Si P est infini, notons $k_0 < k_1 < \cdots < k_j < \cdots$ ses éléments. Par construction

$$x_{k_0} \geq x_{k_1} \geq \cdots \geq x_{k_j} \geq \cdots$$

et donc la sous-suite de nombres réels $(x_{k_j})_j$ est décroissante. Etant de plus minorée par hypothèse, elle est convergente dans $(\mathbb{R}, |\cdot|)$.
Si P est fini, il existe k_0 strictement plus grand que les entiers de P. Comme $k_0 \notin P$, il existe $k_1 > k_0$ tel que $x_{k_1} > x_{k_0}$. Or $k_1 \notin P$, donc il existe $k_2 > k_1$ tel que $x_{k_2} > x_{k_1}$. En continuant ainsi de proche en proche, on peut construire une sous-suite de nombres réels $(x_{k_j})_j$ croissante. Comme de plus elle est majorée, elle est convergente dans $(\mathbb{R}, |\cdot|)$.

2 - On en déduit la démonstration dans le cas général.
Les n suites de nombres réels $(x_{ki})_k$ sont alors toutes bornées pour $i = 1, \ldots, n$.
Par le cas précédent dans $\mathbb{R}$, on peut extraire de la suite donnée $(x_k)_k$ une sous-suite $(x_{1_k})_k$ telle que les premières composantes des points qui la composent forment une suite $(x_{1_k1})_k$ convergeant vers un a_1 dans $(\mathbb{R}, |\cdot|)$.
De cette sous-suite $(x_{1_k})_k$ on peut de nouveau en extraire une sous-suite $(x_{2_k})_k$ telle que les deuxièmes composantes des points qui la composent forment une suite $(x_{2_k2})_k$ convergeant vers un a_2 dans $(\mathbb{R}, |\cdot|)$; de plus $(x_{2_k1})_k$ convergeant vers a_1 dans $(\mathbb{R}, |\cdot|)$ comme sous-suite de la suite convergente $(x_{1_k1})_k$.
Ainsi de suite. Après n étapes, on obtient une sous-suite $(x_{n_k})_k$ de la suite initiale $(x_k)_k$ telle que les suites des composantes $(x_{n_ki})_k$ convergent vers un a_i dans $(\mathbb{R}, |\cdot|)$ pour chaque $i = 1, \ldots, n$.
Par conséquent la sous-suite $(x_{n_k})_k$ converge vers le vecteur $(a_1, \ldots, a_n)$ dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|_\infty)$. $\diamond$

On en déduit en particulier le résultat important suivant :

Corollaire 19. *Toutes les normes sur* $\mathbb{R}^n$ *sont équivalentes.*

Démonstration. Il suffit de comparer une norme $\|\cdot\|$ quelconque sur $\mathbb{R}^n$ à la norme particulière $\|\cdot\|_\infty$.

1 - Pour tout $x \in \mathbb{R}^n$ on a

$$\|x\| = \|\sum_{i=1}^n x_i e_i\| \leq \sum_{i=1}^n |x_i|\,\|e_i\| \leq \sum_{i=1}^n \|e_i\|\,\|x\|_\infty.$$

Donc il existe une constante $C > 0$ telle que pour tout $x \in \mathbb{R}^n$ on a

$$\|x\| \leq C\,\|x\|_\infty.$$

2 - L'ensemble $\{\|x\|;\, x \in \mathbb{R}^n,\, \|x\|_\infty = 1\}$ est un sous-ensemble non vide et minoré dans $\mathbb{R}$, donc il admet un plus grand minorant m, appelé borne inférieure et noté

$$m = \inf\{\|x\|;\, x \in \mathbb{R}^n,\, \|x\|_\infty = 1\} \quad \text{ou} \quad m = \inf_{\substack{x\in\mathbb{R}^n \\ \|x\|_\infty = 1}} \|x\|.$$

Ainsi pour tout $x \in \mathbb{R}^n$ vérifiant $\|x\|_\infty = 1$ on a

$$m \leq \|x\|.$$

Par conséquent, pour tout $x \in \mathbb{R}^n$ avec $x \neq 0$ on a

$$m \leq \left\|\frac{x}{\|x\|_\infty}\right\|,$$

c'est-à-dire

$$m\,\|x\|_\infty \leq \|x\|$$

par homogénéité de la norme, ce résultat étant également valable pour $x = 0$.

Il reste à montrer que $m > 0$. Or, par définition de la borne inférieure, il existe une suite $(x_k)_k$, dite suite minimisante, telle que

$$\|x_k\|_\infty = 1 \quad \text{et} \quad \lim_{k\to\infty} \|x_k\| = m$$

Comme la suite $(x_k)_k$ est bornée dans $(\mathbb{R}^n, \|\cdot\|_\infty)$, elle contient une sous-suite $(x_{k_j})_j$ qui converge vers un a dans $(\mathbb{R}^n, \|\cdot\|_\infty)$ d'après le théorème **??** de Bolzano-Weierstrass.

Comme de plus, par inégalité triangulaire,

$$|\,\|x_{k_j}\|_\infty - \|a\|_\infty\,| \leq \|x_{k_j} - a\|_\infty$$

alors la suite $(\|x_{k_j}\|_\infty)_j$ converge vers $\|a\|_\infty$ dans $\mathbb{R}$, et donc en particulier $\|a\|_\infty = 1$.

Comme enfin

$$|\,\|x_{k_j}\| - \|a\|\,| \leq \|x_{k_j} - a\| \leq C\,\|x_{k_j} - a\|_\infty$$

on en déduit que la suite $(\|x_{k_j}\|)_j$ converge vers $\|a\|$ dans $\mathbb{R}$.

Or cette sous-suite converge aussi vers m car la suite $(\|x_k\|)_k$ converge vers m dans $\mathbb{R}$. Donc par l'unicité de la limite on a

$$\|a\| = m$$

avec de plus $m > 0$ car $a \neq 0$ puisque $\|a\|_\infty = 1$. D'où la conclusion. $\diamond$

Il s'ensuit en particulier qu'une suite bornée dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$ est bornée dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|_*)$ pour toute autre norme $\|\cdot\|_*$ sur $\mathbb{R}^n$; on dit alors que la suite est bornée dans $\mathbb{R}^n$.

2.2. Voisinage, ouvert, fermé, borné de $\mathbb{R}^n$

Définition 20. *Soit $\|\cdot\|$ une norme sur $\mathbb{R}^n$, a un point de $\mathbb{R}^n$ et r un réel > 0. Dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$*

- la boule ouverte de centre a et de rayon r est l'ensemble

$$B(a,r) = \{x \in \mathbb{R}^n;\ \|x-a\| < r\},$$

- la boule fermée de centre a et de rayon r est l'ensemble

$$\overline{B}(a,r) = \{x \in \mathbb{R}^n;\ \|x-a\| \leq r\},$$

- la sphère de centre a et de rayon r est l'ensemble

$$S(a,r) = \{x \in \mathbb{R}^n;\ \|x-a\| = r\}.$$

Exemple. Dans l'e.v.n. $(\mathbb{R}, |\cdot|)$ muni de la valeur absolue

- la boule ouverte $B(a,r)$ est l'intervalle ouvert d'extrémités $a-r$ et $a+r$, noté $]a-r, a+r[$ et défini par

$$]a-r, a+r[= \{x \in \mathbb{R};\ a-r < x < a+r\},$$

- la boule fermée $\overline{B}(a,r)$ est l'intervalle fermé d'extrémités $a-r$ et $a+r$, noté $[a-r, a+r]$ et défini par

$$[a-r, a+r] = \{x \in \mathbb{R};\ a-r \leq x \leq a+r\},$$

- la sphère $S(a,r)$ est constituée des deux extrémités $a-r$ et $a+r$.

Définition 21. *Soit U sous-ensemble de $\mathbb{R}^n$ et a un point de U. Les propositions suivantes sont équivalentes :*

1 - il existe une norme $\|\cdot\|$ de $\mathbb{R}^n$ telle que U contienne une boule (ouverte ou fermée) de centre a de l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$,

2 - pour toute norme $\|\cdot\|$ de $\mathbb{R}^n$, U contient une boule (ouverte ou fermée) de centre a de l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$.

On dit alors que U est un voisinage de a dans $\mathbb{R}^n$.

Démonstration. L'équivalence entre deux normes $\|\cdot\|$ et $\|\cdot\|^*$ montre qu'une boule de centre a dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$ contient une boule de centre a dans l'e.v.n. $(\mathbb{R}^n, \|\cdot\|^*)$, et inversement. Ce qui permet de conclure à l'équivalence des deux propositions. $\diamond$

Définition 22. *Soit U un sous-ensemble de $\mathbb{R}^n$. On dit que*

- U est ouvert dans $\mathbb{R}^n$ si U est vide ou voisinage de chacun de ses points,

- U est fermé dans $\mathbb{R}^n$ si son complémentaire est ouvert.

De même :

Définition 23. *Soit U sous-ensemble de $\mathbb{R}^n$. Les propositions suivantes sont équivalentes :*
1 - il existe une norme $\|\cdot\|$ de $\mathbb{R}^n$ telle que U soit contenu dans une boule (ouverte ou fermée) de l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$,
2 - pour toute norme $\|\cdot\|$ de $\mathbb{R}^n$, U est contenu dans une boule (ouverte ou fermée) de l'e.v.n. $(\mathbb{R}^n, \|\cdot\|)$.
On dit alors que U est borné dans $\mathbb{R}^n$.

Ainsi la notion de voisinage, et donc les notions d'ouvert et de fermé, et la notion de borné ne dépendent pas d'une norme particulière alors que la notion de boule (qui sert à introduire ces notions) est définie à partir d'une norme particulière.

Exemple. Dans $\mathbb{R}^n$, une boule ouverte $B(a,r)$ (définie à partir d'une norme) est un ouvert borné, la boule fermée $\overline{B}(a,r)$ et la sphère $S(a,r)$ sont fermées bornées.

Exemple. Il existe des sous-ensembles qui ne sont ni ouverts, ni fermés. Ainsi si a et b sont deux nombres réels tels que $a < b$, le sous-ensemble de $\mathbb{R}$

$$[a,b[= \{x \in \mathbb{R};\ a \leq x < b\}$$

n'est pas ouvert dans $\mathbb{R}$ (car le point a n'a pas de voisinage dans $\mathbb{R}$ contenu dans $[a,b[$) et n'est pas fermé car son complémentaire n'est pas ouvert (car le point b n'a pas de voisinage dans $\mathbb{R}$ contenu dans ce complémentaire). Cet ensemble est appelé un intervalle semi-ouvert d'extrémités a et b.

Exemple. Dans $\mathbb{R}^n$
1 - l'ensemble vide et $\mathbb{R}^n$ sont des ouverts,
2 - une union quelconque de sous-ensembles ouverts est un ouvert,
3 - une intersection finie de sous-ensembles ouverts est un ouvert.

Exemple. Dans $\mathbb{R}^n$
1 - l'ensemble vide et $\mathbb{R}^n$ sont des fermés,
2 - une intersection quelconque de sous-ensembles fermés est un fermé,
3 - une union finie de sous-ensembles fermés est un fermé.

Exemple. Dans $\mathbb{R}^n$
1 - l'ensemble vide est borné et $\mathbb{R}^n$ n'est pas borné,
2 - une intersection quelconque de sous-ensembles bornés est bornée,
3 - une union finie de sous-ensembles bornés est bornée.

On peut réécrire la définition de la convergence d'une suite dans $\mathbb{R}^n$.

Définition 24. *Soit* $(x_k)_k$ *une suite de* $\mathbb{R}^n$*. On dit que la suite* $(x_k)_k$ *est convergente dans* $\mathbb{R}^n$ *s'il existe* x *appartenant à* $\mathbb{R}^n$ *tel que pour tout voisinage* V *de* x *dans* $\mathbb{R}^n$ *il existe un entier* $K > 0$ *tel que* $x_k \in V$ *pour* $k \geq K$*.*

Cet élément x *est alors unique. On dit que la suite* $(x_k)_k$ *converge vers* x *dans* $\mathbb{R}^n$*, que* x *est la limite de la suite* $(x_k)_k$ *et on note*

$$\lim_{k\to\infty} x_k = x.$$

On peut traduire cette définition dans un e.v.n. $(\mathbb{R}^n, \|\cdot\|)$. D'après les définitions on a immédiatement :

Proposition 25. *Soit* $(x_k)_k$ *une suite de* $\mathbb{R}^n$*,* x *un point de* $\mathbb{R}^n$ *et* $\|\cdot\|$ *une norme de* $\mathbb{R}^n$*. Alors la suite* $(x_k)_k$ *converge vers* x *dans* $\mathbb{R}^n$ *si et seulement si elle converge vers* x *dans l'e.v.n.* $(\mathbb{R}^n, \|\cdot\|)$*.*

Du théorème **??** démontré pour la norme $\|\cdot\|_\infty$ on déduit alors le théorème général suivant

Corollaire 26 (Théorème de Bolzano-Weierstrass). *Toute suite de points de* $\mathbb{R}^n$ *bornée dans* $\mathbb{R}^n$ *contient une sous-suite convergeant dans* $\mathbb{R}^n$*.*

Comme dans $\mathbb{R}$, le critère de Cauchy caractérise les suites convergentes :

Définition 27. *On dit que la suite* $(x_k)_k$ *de* $\mathbb{R}^n$ *est une suite de Cauchy si pour tout voisinage* V *de* 0 *dans* $\mathbb{R}^n$ *il existe un entier* $K > 0$ *tel que* $x_k - x_l \in V$ *pour* $k, l \geq K$*.*

Théorème 28 (de Cauchy). *Une suite* $(x_k)_k$ *de* $\mathbb{R}^n$ *est convergente dans* $\mathbb{R}^n$ *si et seulement si elle est une suite de Cauchy.*

Démonstration. En prenant par exemple la norme $\|\cdot\|_\infty$ sur $\mathbb{R}^n$ on se ramène à des suites dans $\mathbb{R}$ dans lequel on a le critère de Cauchy. $\diamond$

Proposition 29. *Soit* (x_k) *une suite convergente vers* x *dans* $\mathbb{R}^n$*. Alors*
1 - toute sous-suite $(x_{k_j})_j$ *de la suite* (x_k) *converge vers* x *dans* $\mathbb{R}^n$*,*
2 - la suite $(x_k)_k$ *est bornée dans* $\mathbb{R}^n$*.*

Démonstration. On démontre par exemple la seconde conclusion. Soit $\|\cdot\|$ une norme de $\mathbb{R}^n$.

Etant donné $\varepsilon > 0$ il existe un entier $K > 0$ tel que $\|x_k - x\| \leq \varepsilon$ pour tout $k \geq K$, donc $\|x_k\| \leq \varepsilon + \|x\|$ pour tout $k \geq K$.

Posant $r = \max(\|x_0\|, \ldots, \|x_K\|, \varepsilon + \|x\|)$ on a alors $\|x_k\| \leq r$ pour tout k, ce qui signifie que la suite $(x_k)_k$ est bornée dans $\mathbb{R}^n$. $\diamond$

Théorème 30. *Un sous-ensemble* U *de* $\mathbb{R}^n$ *est fermé si et seulement si la limite de toute suite de points de* U *convergente dans* $\mathbb{R}^n$ *appartient à* U*.*

Démonstration. 1 - Supposons U fermé dans $\mathbb{R}^n$ et soit $(x_k)_k$ une suite de points de U convergeant vers x dans $\mathbb{R}^n$.

Si cette limite x n'appartient pas à U, il existe un voisinage de x qui ne rencontre pas U puisque le complémentaire de U est ouvert. Ceci est contraire à la convergence de la suite $(x_k)_k$ vers x, donc $x \in U$.

2 - Inversement supposons U non fermé dans $\mathbb{R}^n$. Son complémentaire n'étant pas ouvert, il existe un point x de ce complémentaire tel que tout voisinage de x rencontre U. En particulier pour une norme $\|\cdot\|$ de $\mathbb{R}^n$, la boule $B(x, r)$ rencontre U pour tout $r > 0$ et donc pour tout entier $k \geq 1$, il existe $x_k \in U$ tel que

$$\|x_k - x\| \leq \frac{1}{k}\cdot$$

Ainsi si U n'est pas fermé, il existe une suite (x_k) de points de U convergeant vers un point x non situé dans U. $\diamond$

Corollaire 31. *Un sous-ensemble U de $\mathbb{R}^n$ est fermé et borné dans $\mathbb{R}^n$ si et seulement si toute suite de points de U contient une sous-suite convergeant dans $\mathbb{R}^n$ vers un point de U.*

Démonstration. 1 - On suppose que toute suite de points de U contient une sous-suite convergeant dans $\mathbb{R}^n$ vers un point de U.

Tout d'abord U est fermé dans $\mathbb{R}^n$. Soit en effet (x_k) une suite de points de U convergeant vers un point x dans $\mathbb{R}^n$. D'une part, par l'hypothèse, cette suite contient une sous-suite qui converge vers un point de U, et d'autre part, par la proposition **??**, toute sous-suite converge aussi vers x. On en déduit donc que $x \in U$ par unicité de la limite, puis que U est fermé par le théorème **??**.

Ensuite si U n'est pas borné dans $\mathbb{R}^n$, et si $\|\cdot\|$ est une norme sur $\mathbb{R}^n$ il existe une suite (x_k) de points de U tels que

$$\|x_{k+1}\| \geq \|x_k\| + 1.$$

Or par la proposition **??**, toute suite convergente étant nécessairement bornée, cette suite ne contient donc pas de sous-suite convergente, ce qui est contraire à l'hypothèse.

2 - La réciproque est une conséquence du théorème de Bolzano-Weierstrass et du théorème précédent. $\diamond$

Remarque. On peut donner une autre caractérisation des sous-ensembles fermés bornés de $\mathbb{R}^n$.

On dit qu'un sous-ensemble U de $\mathbb{R}^n$ est *compact* si, de toute famille $(U_i)_{i\in I}$ telle que U_i est un ouvert de $\mathbb{R}^n$ pour tout $i \in I$ et $U \subset \cup_{i\in I} U_i$, on peut extraire une famille finie $(U_j)_{j\in J}$ avec J sous-ensemble fini de I telle que $U \subset \cup_{j\in J} U_j$.

Avec cette définition on a alors le théorème de Borel-Lebesgue : U est fermé et borné si et seulement si U est compact.

2.3. Adhérence, intérieur

Proposition 32. *Soit U un sous-ensemble de $\mathbb{R}^n$.*

- La réunion de tous les ouverts de $\mathbb{R}^n$ contenu dans U est le plus grand ouvert de $\mathbb{R}^n$ contenu dans U. Il est noté $\overset{\circ}{U}$ et appelé intérieur de U. Il peut être vide.

- L'intersection de tous les fermés de $\mathbb{R}^n$ contenant U est le plus petit fermé de $\mathbb{R}^n$ contenant U. Il est noté $\overline{U}$ et appelé adhérence de U. Il peut être égal à $\mathbb{R}^n$.

Proposition 33. *Un sous-ensemble de $\mathbb{R}^n$ est ouvert (resp. fermé) si et seulement si il est égal à son intérieur (resp. adhérence).*

Proposition 34. *Soit U un sous-ensemble de $\mathbb{R}^n$ et x un point de U. Les trois propositions suivantes sont équivalentes :*

1 - x appartient à $\overline{U}$,

2 - tout voisinage de x contient un point de U,

3 - x est limite d'une suite de points de U.

On dit alors que x est adhérent à U.

L'adhérence de U est donc égal à l'ensemble des points adhérents à U.

Exemples. Les points de la la sphère $S(a, r)$ sont adhérents à la boule ouverte $B(a, r)$.

L'adhérence de la boule ouverte $B(a, r)$ est la boule fermée $\overline{B}(a, r)$, d'où la notation.

Chapitre 2 : Fonctions continues

Les espaces $\mathbb{R}^n$ et $\mathbb{R}^p$ étant munis des structures topologiques introduites dans le chapitre 1, on s'intéresse dans ce chapitre à des fonctions f définies dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$

$$f : U \subset \mathbb{R}^n \to \mathbb{R}^p : x \mapsto f(x).$$

Les composantes du vecteur $f(x)$ de $\mathbb{R}^p$ sont notées $f_j(x)$ pour $j = 1, \ldots, p$, c'est-à-dire $f(x) = (f_1(x), \ldots, f_p(x))$, et ainsi à la fonction f on associe les fonctions f_j définies dans U à valeurs dans $\mathbb{R}$ par

$$f_j : U \subset \mathbb{R}^n \to \mathbb{R} : x \mapsto f_j(x)$$

pour $j = 1, \ldots, p$.

1. Fonctions continues en un point

Définition 35. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. On dit que f est continue en a si pour tout voisinage W de $f(a)$ dans $\mathbb{R}^p$, il existe un voisinage V de a dans $\mathbb{R}^n$ tel que $f(x) \in W$ pour tout $x \in V \cap U$.*

Le voisinage V dépend du voisinage W.

Ceci signifie que $f(x)$ est « arbitrairement près » de $f(a)$ dès que x est « suffisamment près » de a dans U, et que $f(x)$ « tend vers » $f(a)$ quand x « tend vers » a dans U. On le note $f(x) \to f(a)$ quand $x \to a$ dans U (ou simplement $x \to a$) et on écrit $\lim\limits_{x \to a,\, x \in U} f(x) = f(a)$ (ou simplement $\lim\limits_{x \to a} f(x) = f(a)$).

Remarque. La continuité de f en un point a est une notion locale en ce sens qu'il suffit de connaître f dans $U' \cap U$ pour un voisinage U' de a dans $\mathbb{R}^n$ aussi petit soit-il.

Proposition 36. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, a un point de U, $\|\cdot\|_n$ une norme de $\mathbb{R}^n$ et $\|\cdot\|_p$ une norme de $\mathbb{R}^p$. Alors la fonction f est continue en a si et seulement si pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour tout $x \in U$ vérifiant $\|x - a\|_n \le \delta$ on ait $\|f(x) - f(a)\|_p \le \varepsilon$.*

La continuité peut donc s'exprimer avec n'importe quelle norme.
Le nombre δ dépend de ε ainsi que des normes $\|\cdot\|_n$ et $\|\cdot\|_p$.
On peut remplacer les inégalités larges $\le$ par des inégalités strictes $<$.

Démonstration. Ces caractérisations résultent immédiatement de la définition d'un voisinage de $f(a)$.
1 - On suppose que la fonction f est continue en a.

De façon générale on note $B_m(x,r)$ la boule de centre x et de rayon r dans $\mathbb{R}^m$ pour la norme $\|\cdot\|_m$ pour $m=n$ et $m=p$.

Soit $\varepsilon>0$ donné. La boule $B_p(f(a),\varepsilon)$ étant un voisinage de $f(a)$ dans $\mathbb{R}^p$, il existe un voisinage V de a dans $\mathbb{R}^n$ tel que $f(x)\in B_p(f(a),\varepsilon)$ pour tout $x\in V\cap U$.

Or par définition d'un voisinage, il existe un $\delta>0$ tel que $B_n(a,\delta)\subset V$.

Ainsi on en déduit que pour tout $\varepsilon>0$, il existe un $\delta>0$ tel que pour tout $x\in U$ vérifiant $\|x-a\|_n\le\delta$ on ait $\|f(x)-f(a)\|_p\le\varepsilon$.

2 - La réciproque se montre de la même façon. $\diamond$

Exemple. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^n$. Alors la fonction $\|\cdot\|$ définie dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ est continue en chaque point a de $\mathbb{R}^n$ car pour tout $x\in\mathbb{R}^n$ on a

$$|\,\|x\|-\|a\|\,|\le\|x-a\|.$$

Il peut être pratique de se ramener à l'utilisation de suites.

Proposition 37. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Alors la fonction f est continue en a si et seulement si pour toute suite $(x_k)_k$ de points de U convergeant vers a dans $\mathbb{R}^n$, la suite $(f(x_k))_k$ converge vers $f(a)$ dans $\mathbb{R}^p$.*

Démonstration. On fixe une norme $\|\cdot\|_n$ sur $\mathbb{R}^n$ et une norme $\|\cdot\|_p$ sur $\mathbb{R}^p$.

1 - On suppose que la fonction f est continue en a, et soit $(x_k)_k$ une suite de points de U convergeant vers a.

Soit $\varepsilon>0$ fixé. Comme f est continue en a, il existe un $\delta>0$ tel que pour tout $x\in U$ vérifiant $\|x-a\|_n\le\delta$ on ait $\|f(x)-f(a)\|_p\le\varepsilon$.

Comme la suite $(x_k)_k$ converge vers a, à ce $\delta>0$ on peut associer un K tel que $\|x_k-a\|_n\le\delta$ pour $k\ge K$.

Ainsi pour tout $\varepsilon>0$, il existe un K tel que $\|f(x_k)-f(a)\|_p\le\varepsilon$ pour $k\ge K$, ce qui signifie que la suite $(f(x_k))_k$ converge vers $f(a)$ dans $\mathbb{R}^p$.

2 - Inversement on suppose que f n'est pas continue en a. Donc il existe un $\varepsilon>0$ tel que pour tout $\delta>0$ il existe un $x\in U$ vérifiant $\|x-a\|_n\le\delta$ et $\|f(x)-f(a)\|_p\ge\varepsilon$.

En particulier en prenant δ de la forme $\dfrac{1}{k}$ pour k entier, il existe ainsi une suite $(x_k)_k$ de points de U convergeant vers a et telle que $\|f(x_k)-f(a)\|_p\ge\varepsilon$ pour tout k, c'est-à-dire telle que la suite $(f(x_k))_k$ ne converge pas vers $f(a)$ dans $\mathbb{R}^p$. $\diamond$

Proposition 38. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Alors f est continue en a si et seulement si chaque composante f_j de f est continue en a pour $j=1,\dots,p$.*

Démonstration. On fixe une norme quelconque $\|\cdot\|_n$ sur $\mathbb{R}^n$ et la norme $\|\cdot\|_{p\infty}$ sur $\mathbb{R}^p$ où pour $x \in \mathbb{R}^p$ on note

$$\|x\|_{p\infty} = \sup_{1\leq i\leq p} |x_i|.$$

1 - On suppose que f est continue en a.

D'après la proposition **??**, pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour tout $x \in U$ vérifiant $\|x-a\|_n \leq \delta$ on ait $\|f(x)-f(a)\|_{p\infty} \leq \varepsilon$, donc $|f_i(x)-f_i(a)| \leq \varepsilon$ pour $i = 1,\ldots,p$.

Cette même proposition **??** implique alors que les fonctions f_i sont continues en a pour $i = 1,\ldots,p$.

2 - On montre la réciproque de la même façon. ◇

Proposition 39. *Soit f et g deux fonctions définies dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, h une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U.*

1 - Si les fonctions f et g sont continues en a, la fonction $f+g$ est continue en a,

2 - Si les fonctions h et f sont continues en a, la fonction hf est continue en a,

3 - Si la fonction h est continue en a et non nulle en a, alors il existe un voisinage V de a tel que $h(x) \neq 0$ pour $x \in V \cap U$ et la fonction $\dfrac{1}{h}$ est continue en a.

Démonstration. On démontre par exemple la propriété 3 en munissant $\mathbb{R}^n$ d'une norme $\|\cdot\|_n$ et $\mathbb{R}$ de la valeur absolue.

Comme h est continue en a, pour tout $\varepsilon > 0$, il existe un $\delta(\varepsilon) > 0$ tel que pour tout $x \in U$ vérifiant $\|x-a\|_n \leq \delta(\varepsilon)$ on ait $|h(x)-h(a)| \leq \varepsilon$.

Comme $h(a) \neq 0$, en prenant en particulier $\varepsilon = \dfrac{|h(a)|}{2}$, pour tout $x \in U$ vérifiant $\|x-a\|_n \leq \delta\Big(\dfrac{|h(a)|}{2}\Big)$ on a $|h(x)-h(a)| \leq \dfrac{|h(a)|}{2}$ et donc $|h(x)| \geq \dfrac{|h(a)|}{2}$.

Ainsi pour tout $x \in U$ vérifiant $\|x-a\|_n \leq \inf\Big(\delta(\varepsilon), \delta\Big(\dfrac{|h(a)|}{2}\Big)\Big)$ on a $h(x) \neq 0$ et

$$\Big|\frac{1}{h(x)} - \frac{1}{h(a)}\Big| = \frac{|h(x)-h(a)|}{|h(x)|\,|h(a)|} \leq \frac{2}{|h(a)|^2}\,|h(x)-h(a)| \leq \frac{2}{|h(a)|^2}\,\varepsilon,$$

ce qui permet de conclure. ◇

Proposition 40. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et g une fonction définie dans un sous-ensemble V de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^q$ tel que $f(U) \subset V$. Si f est continue en un point a de U et si g est continue en $f(a)$, alors la fonction $g \circ f$ est continue en a.*

Démonstration. On démontre la propriété avec des normes $\|\cdot\|_m$ sur $\mathbb{R}^m$ pour $m = n$, p et q.

Comme g est continue en $f(a)$, pour tout $\varepsilon > 0$, il existe un $\delta_p > 0$ tel que pour tout $y \in V$ vérifiant $\|y-f(a)\|_p \leq \delta_p$ on ait $\|g(y)-g(f(a))\|_q \leq \varepsilon$.

Comme f est continue en a, au nombre $\delta_p > 0$, on peut associer un $\delta_n > 0$ tel que pour tout $x \in U$ vérifiant $\|x - a\|_n \leq \delta_n$ on ait $\|f(x)) - f(a)\|_p \leq \delta_p$.

Comme de plus $f(U) \subset V$, on en déduit que pour tout $\varepsilon > 0$, il existe un $\delta_n > 0$ tel que pour tout $x \in U$ vérifiant $\|x - a\|_n \leq \delta_n$ on ait $\|g(f(x)) - g(f(a))\|_q \leq \varepsilon$. ◇

Proposition 41. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, V un sous-ensemble de U et a un point de V. Si la fonction f est continue en a, alors sa restriction $f_{|V}$ à V est continue en a.*

La restriction $f_{|V}$ à V est la fonction définie dans V telle que $f_{|V}(x) = f(x)$ pour tout $x \in V$.

Démonstration. La démonstration est évidente. ◇

Mais la réciproque n'est pas toujours vraie suivant la structure de V.

Exemple. Soit f la fonction définie dans $\mathbb{R}$ par

$$\begin{array}{lll} f(x) = 1 & \text{pour} & x \geq 0 \\ f(x) = 0 & \text{pour} & x < 0 \end{array}$$

et soit $V = \{x \in \mathbb{R};\, x \geq 0\}$.

Alors $f_{|V}$ est continue en 0 mais f n'est pas continue en 0.

Plus particulièrement

Proposition 42. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Soit $U_i = \{x_i \in \mathbb{R};\, (a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n) \in U\}$ pour $i = 1, \ldots, n$ (avec des conventions d'écriture évidentes pour $i = 1$ et $i = n$).*

Si la fonction f est continue en a, l'application partielle

$$f(a_1, \ldots, a_{i-1}, \cdot, a_{i+1}, \ldots, a_n) : U_i \subset \mathbb{R} \to \mathbb{R}^p : x_i \mapsto f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n)$$

est continue au point a_i pour $i = 1, \ldots, n$.

Démonstration. La démonstration est évidente. ◇

Mais la réciproque n'est pas toujours vraie.

Exemple. On considère la fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{array}{llll} f(x, y) = & \dfrac{xy}{x^2 + y^2} & \text{si} & (x, y) \neq (0, 0) \\ f(x, y) = & 0 & \text{si} & (x, y) = (0, 0). \end{array}$$

La fonction partielle $f(\cdot, 0) : \mathbb{R} \to \mathbb{R} : x \mapsto f(x, 0)$ obtenue en fixant $y = 0$ et la fonction partielle $f(0, \cdot) : \mathbb{R} \to \mathbb{R} : y \mapsto f(0, y)$ obtenue en fixant $x = 0$ sont des fonctions constantes (égales à 0) dans $\mathbb{R}$, donc continues en 0. Cependant la fonction f n'est pas continue en $(0, 0)$.

En effet si f est continue en $(0,0)$, étant donné un $\varepsilon < \dfrac{1}{2}$ il existe un $\delta > 0$ tel que pour $\|(x,y)\| \leq \delta$ on ait $|f(x,y)| = |f(x,y) - f(0,0)| \leq \varepsilon$.

Or pour tout $x \neq 0$ dans $\mathbb{R}$ on a $f(x,x) = \dfrac{1}{2}$, ce qui contredit l'estimation précédente pour x tel que $\|(x,x)\| \leq \delta$.

Exemple. De même la fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{array}{ll} f(x,y) = 1 & \text{si} \quad x = 0 \ \text{ ou } \ y = 0, \\ f(x,y) = 0 & \text{sinon} \end{array}$$

n'est pas continue en $(0,0)$ alors que les applications partielles $f(\cdot,0)$ et $f(0,\cdot)$ sont continues en 0.

2. Fonctions continues dans un ensemble

Définition 43. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et V un sous-ensemble de U. On dit que f est continue dans V si f est continue en chaque point de V.*

Remarque. Si f est continue dans V, la fonction $f_{|V}$ restriction de f à V est continue dans V, mais inversement $f_{|V}$ peut être continue dans V sans que f soit continue dans V.

Exemple. Soit f la fonction définie dans $\mathbb{R}$ par

$$\begin{array}{lll} f(x) = 1 & \text{pour} & x \geq 0 \\ f(x) = 0 & \text{pour} & x < 0 \end{array}$$

et soit $V = \{x \in \mathbb{R};\, x \geq 0\}$.

Alors $f_{|V}$ est continue dans V mais f n'est pas continue dans V car non continue en 0.

Proposition 44. *Soit f une fonction définie dans un sous-ensemble ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$. Si f est continue dans U et si V est un sous-ensemble ouvert dans $\mathbb{R}^p$, alors $f^{-1}(V)$ est ouvert dans $\mathbb{R}^n$.*

Démonstration. Soit a un point quelconque de $f^{-1}(V)(= \{x \in U;\, f(x) \in V\})$.

Puisque V est ouvert et $f(a) \in V$, il existe une boule $B_p(f(a),\varepsilon)$ de $\mathbb{R}^p$ telle que

$$B_p(f(a),\varepsilon) \subset V.$$

Puisque U est ouvert et f est continue en a, il existe une boule $B_n(a,\delta)$ de $\mathbb{R}^n$ telle que

$$B_n(a,\delta) \subset U \quad \text{et} \quad f(B_n(a,\delta)) \subset B_p(f(a),\varepsilon).$$

Par conséquent

$$B_n(a,\delta) \subset f^{-1}(V),$$

donc $f^{-1}(V)$ est un voisinage de a. $\diamond$

Théorème 45. *Soit f une fonction définie dans un sous-ensemble U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$. Si f est continue dans U et si U est fermé et borné dans $\mathbb{R}^n$, alors $f(U)$ est un sous-ensemble fermé et borné de $\mathbb{R}^p$.*

Démonstration. On utilise la caractérisation par les suites des ensembles fermés bornés donnée dans le chapitre 1.

Soit $(y_k)_k$ une suite de points de $f(U)$. Il existe donc une suite $(x_k)_k$ de points de U tels que $y_k = f(x_k)$ pour tout k.

Puisque U est fermé et borné, il existe une sous-suite $(x_{k_j})_j$ de la suite $(x_k)_k$ qui converge vers un point x de U. Comme f est continue dans U, donc en particulier en x, la suite $(f(x_{k_j}))_j$ converge vers $f(x)$.

Ainsi la suite $(y_k)_k$ de $f(U)$ contient une sous-suite $(y_{k_j})_j$ qui converge vers le point $f(x)$ de $f(U)$.

Par conséquent $f(U)$ est un sous-ensemble fermé et borné de $\mathbb{R}^p$. ◇

En particulier si la fonction f est à valeurs réelles on donne un résultat important pour les problèmes d'optimisation.

Corollaire 46 (Théorème de Weierstrass). *Soit f une fonction définie dans un sous-ensemble non vide U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$. Si f est continue dans U et si U est fermé et borné dans $\mathbb{R}^n$, alors*

1 - $f(U)$ est un sous-ensemble borné de $\mathbb{R}$,

2 - il existe x_m et x_M appartenant à U tels que

$$f(x_m) = \inf\{f(x);\, x \in U\} \quad \textit{et} \quad f(x_M) = \sup\{f(x);\, x \in U\}.$$

On dit aussi que f est bornée dans U et atteint ses bornes dans U. Ainsi f admet un minimum global et un maximum global dans U (ces notions seront reprises dans le chapitre 6).

Démonstration. 1 - D'après le théorème **??**, $f(U)$ est un sous-ensemble non vide, fermé et borné, c'est-à-dire majoré et minoré, dans $\mathbb{R}$.

2 - Le sous-ensemble $f(U)$ étant non vide et minoré dans $\mathbb{R}$, il admet une borne inférieure m (c'est-à-dire un plus grand minorant) notée $m = \inf\{f(x);\, x \in U\}$ ou $m = \inf\limits_{x\in U} f(x)$. De plus $m \in f(U)$: il existe en effet une suite $(y_k)_k$ de $f(U)$ convergeant vers m; comme $f(U)$ est fermé, la limite m appartient à $f(U)$.

Ainsi il existe x_m appartenant à U tel que

$$f(x_m) = \inf\{f(x);\, x \in U\}.$$

On raisonne de même pour la borne supérieure. ◇

Pour étudier le cas où U est non borné on doit faire une hypothèse sur la fonction f.

Définition 47. *Une fonction f définie dans un sous-ensemble non borné U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ est dite cœrcive si pour une norme $\|\cdot\|$ dans $\mathbb{R}^n$*

$$\lim_{\substack{\|x\|\to+\infty \\ x\in U}} f(x) = +\infty.$$

Ceci signifie que pour tout $M > 0$, il existe $r > 0$ tel que pour tout $x \in U$ vérifiant $\|x\| \geq r$, on a $f(x) \geq M$.

Corollaire 48. *Soit U un sous-ensemble fermé de $\mathbb{R}^n$ et f une fonction définie et continue dans U à valeurs dans $\mathbb{R}$, de plus coercive lorsque U n'est pas borné. Alors*
1 - $f(U)$ est minoré dans $\mathbb{R}$,
2 - il existe $x_m \in U$ tel que $f(x_m) = \inf\{f(x);\, x \in U\}$.

Ainsi f admet un minimum global dans U.

De même si $-f$ est coercive lorsque U n'est pas borné, alors f admet un maximum global dans U.

Démonstration. Lorsque U est borné, le résultat est le théorème de Weierstrass.
Lorsque U n'est pas borné, on se ramène au cas borné grâce à la coercivité de f. Pour cela on fixe un point a de U. La coercivité de f entraîne l'existence d'un nombre $r > 0$ tel que $f(a) < f(x)$ pour $x \in U$ vérifiant $\|x\| > r$.
Le sous-ensemble $U_r = \{x \in U;\, \|x\| \leq r\}$ de $\mathbb{R}^n$ étant non vide (car $a \in U_r$), fermé (car intersection de deux fermés) et borné, il existe au moins un $x_m \in U_r$ tel que $f(x_m) = \inf\{f(x);\, x \in U_r\}$ d'après le théorème de Weierstrass.
Or le sous-ensemble $f(U)$ de $\mathbb{R}$ est minoré car pour tout $x \in U$ on a

$$f(x) \geq \inf\{f(x);\, x \in U_r\}$$

d'après le choix de r, donc admet une borne inférieure telle que

$$\inf\{f(x);\, x \in U\} = \inf\{f(x);\, x \in U_r\}. \qquad \diamond$$

3. Applications linéaires, formes bilinéaires, formes quadratiques

Corollaire 49. *Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. Alors*
1 - f est continue dans $\mathbb{R}^n$,
2 - pour toutes normes $\|\cdot\|_n$ et $\|\cdot\|_p$ sur $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement, il existe des constantes $m \geq 0$ et $M \geq 0$ telles que pour tout $x \in \mathbb{R}^n$ on ait

$$m\,\|x\|_n \leq \|f(x)\|_p \leq M\|x\|_n,$$

3 - $f(x) \neq 0$ pour $x \in \mathbb{R}^n$, $x \neq 0$ si et seulement si pour des normes $\|\cdot\|_n$ et $\|\cdot\|_p$ sur $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement, il existe une constante $m > 0$ telle que pour tout $x \in \mathbb{R}^n$ on ait

$$m\,\|x\|_n \leq \|f(x)\|_p.$$

La fonction f étant linéaire, elle est injective dans $\mathbb{R}^n$ si et seulement si $f(x) \neq 0$ pour $x \in \mathbb{R}^n$, $x \neq 0$.

Ainsi la conclusion 3 caractérise l'injectivité de f. Si de plus $p = n$, elle caractérise sa bijectivité, et donc son inversibilité.

Démonstration. 1 - Soit $\|\cdot\|_\infty$ la norme sur $\mathbb{R}^n$ définie par $\|x\|_\infty = \sup_{i=1,\dots,n} |x_i|$ et $\|\cdot\|_p$ une norme sur $\mathbb{R}^p$.

Pour tout $x \in \mathbb{R}^n$ on a

$$\|f(x)\|_p = \|f(\sum_{i=1}^n x_i e_i)\|_p = \|\sum_{i=1}^n x_i\, f(e_i)\|_p \leq \sum_{i=1}^n |x_i|\, \|f(e_i)\|_p \leq \sum_{i=1}^n \|f(e_i)\|_p\, \|x\|_\infty.$$

Donc il existe une constante $C > 0$ telle que pour tout $x \in \mathbb{R}^n$ on a

$$\|f(x)\|_p \leq C\, \|x\|_\infty.$$

D'après la linéarité de f, on en déduit que pour tous x et $a \in \mathbb{R}^n$ on a

$$\|f(x) - f(a)\|_p = \|f(x-a)\|_p \leq C\, \|x-a\|_\infty$$

ce qui assure la continuité de f en tout point a de $\mathbb{R}^n$.

2 - Soit $\|\cdot\|_n$ et $\|\cdot\|_p$ des normes sur $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement.

La sphère $S_n(0,1) = \{x \in \mathbb{R}^n;\ \|x\|_n = 1\}$ étant fermée et bornée et la fonction $\|f\|_p$ étant continue dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, le théorème de Weierstrass assure qu'il existe x_m et $x_M \in S_n(0,1)$ tels que pour tout $x \in S_n(0,1)$

$$\|f(x_m)\|_p \leq \|f(x)\|_p \leq \|f(x_M)\|_p.$$

Par homogénéité de la norme $\|\cdot\|_n$, il en résulte que pour tout $x \in \mathbb{R}^n$

$$m\, \|x\|_n \leq \|f(x)\|_p \leq M\, \|x\|_n$$

avec $m = \|f(x_m)\|_p$ et $M = \|f(x_M)\|_p$.

(On note qu'une majoration du type $\|f(x)\|_p \leq M\, \|x\|_n$ se déduit de l'estimation établie en 1 puisque les normes $\|\cdot\|_n$ et $\|\cdot\|_\infty$ sur $\mathbb{R}^n$ sont équivalentes.)

3 - Avec les notations précédentes, si f est injective, alors en particulier $f(x_m) \neq 0$, donc $m = \|f(x_m)\|_p > 0$.

Inversement s'il existe une constante $m > 0$ telle que pour tout $x \in \mathbb{R}^n$ on ait

$$m\, \|x\|_n \leq \|f(x)\|_p$$

alors f est évidemment injective dans $\mathbb{R}^n$ puisque de plus linéaire. $\diamond$

Remarque. Pour une application linéaire f de $\mathbb{R}^n$ dans $\mathbb{R}^p$, l'ensemble

$$\Big\{ \frac{\|f(x)\|_p}{\|x\|_n};\ x \in \mathbb{R}^n,\ x \neq 0 \Big\}$$

est donc un sous-ensemble non vide et majoré de $\mathbb{R}$. Donc il admet une borne supérieure

$$\|f\|_{np} = \sup\left\{\frac{\|f(x)\|_p}{\|x\|_n};\ x \in \mathbb{R}^n,\ x \neq 0\right\} = \{\|f(x)\|_p;\ x \in \mathbb{R}^n,\ \|x\|_n = 1\}$$

appelée norme subordonnée aux normes $\|\cdot\|_n$ et $\|\cdot\|_p$ sur $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement. D'après le théorème de Weierstrass, il existe $x_M \in \mathbb{R}^n$ avec $\|x_M\|_n = 1$ tel que $\|f\|_{np} = \|f(x_M)\|_p$.

On a des résultats du même type pour les formes bilinéaires et les formes quadratiques. Ainsi par exemple pour les formes quadratiques :

Corollaire 50. *Soit f une forme quadratique sur $\mathbb{R}^n$. Alors*

1 - f est continue dans $\mathbb{R}^n$,

2 - pour toute norme $\|\cdot\|$ sur $\mathbb{R}^n$, il existe des constantes m et M telles que pour tout $x \in \mathbb{R}^n$ on ait

$$m\,\|x\|^2 \leq f(x) \leq M\,\|x\|^2,$$

3 - $f(x) > 0$ pour tout $x \in \mathbb{R}^n$, $x \neq 0$ si et seulement si pour une norme $\|\cdot\|$ sur $\mathbb{R}^n$, il existe une constante $m > 0$ telle que pour tout $x \in \mathbb{R}^n$ on ait

$$m\,\|x\|^2 \leq f(x).$$

Démonstration.

1 - Soit A la matrice (symétrique) représentative de f relativement à la base canonique de $\mathbb{R}^n$.

Pour x et $a \in \mathbb{R}^n$, on a

$$f(x) - f(a) = <Ax, x> - <Aa, a> = <A(x-a), x-a> + 2 <A(x-a), a> .$$

Donc pour la norme euclidienne $\|\cdot\|_2$ sur $\mathbb{R}^n$

$$|f(x) - f(a)| \leq \|A(x-a)\|_2\|x-a\|_2 + 2\|A(x-a)\|_2\|a\|_2$$

et par la proposition précédente, il existe une constante $M \geq 0$ telle que

$$|f(x) - f(a)| \leq M(\|x-a\|_2 + 2\|a\|_2)\|x-a\|_2.$$

Par exemple pour x tel que $\|x-a\|_2 \leq 1$, on a donc

$$|f(x) - f(a)| \leq M(1 + 2\|a\|_2)\|x-a\|_2,$$

d'où l'on déduit la continuité de f en tout point a de $\mathbb{R}^n$.

2 et 3 se vérifient comme dans la démonstration précédente. ◇

Chapitre 3 - Dérivabilité - Différentiabilité

1. Dérivées d'ordre un et deux

On s'intéresse dans ce paragraphe à la dérivation classique par rapport à une variable réelle pour des fonctions de n variables réelles.

1.1. Dérivées. Cas $n = 1$

On examine tout d'abord le cas **n =1**. La notion de dérivée pour les fonctions d'une variable réelle à valeurs réelles dans $\mathbb{R}$ se généralise de façon naturelle pour des fonctions d'une variable réelle à valeurs vectorielles dans $\mathbb{R}^p$ pour $p \geq 1$ sous la forme :

Définition 51. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et a un point de U. On dit que f est une fois dérivable en a si (le quotient différentiel)*

$$\frac{f(a+h)-f(a)}{h}$$

a une limite l dans $\mathbb{R}^p$ quand h tend vers 0 dans $\mathbb{R}$, $h \neq 0$.

Cette limite l, qui est alors unique, est l'élément de $\mathbb{R}^p$ noté $f'(a)$ ou $\dfrac{df}{dx}(a)$ et appelé la dérivée première ou d'ordre un de f en a.

Cette définition signifie que pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour h réel vérifiant $0 < |h| \leq \delta$ on a

$$\left\| \frac{f(a+h)-f(a)}{h} - l \right\| \leq \varepsilon$$

où $\|.\|$ désigne une norme de $\mathbb{R}^p$, et on note

$$\lim_{\substack{h\to 0,\, h\neq 0 \\ h\in\mathbb{R}}} \frac{f(a+h)-f(a)}{h} = l.$$

Cette définition signifie aussi que la fonction ε définie dans un voisinage de 0 dans $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ par

$$\begin{array}{lll} \varepsilon(h) = & \dfrac{f(a+h)-f(a)}{h} - l & \text{pour} \quad h \neq 0 \\ \varepsilon(h) = & 0 & \text{pour} \quad h = 0 \end{array}$$

est continue en 0.

Pour simplifier dans ce qui suit et quand il n'y aura pas d'ambiguité, on dira *dérivable* à la place de *une fois dérivable* et *dérivée* à la place de *dérivée première ou d'ordre un.*

Cette définition peut aussi être donnée de façon équivalente sous la forme suivante :

Définition 52. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et a un point de U. On dit que f est dérivable en a s'il existe un élément l de $\mathbb{R}^p$ tel que*

$$f(a+h) = f(a) + hl + |h|\varepsilon(h)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$, continue en 0 et nulle en 0.

Cet élément l, qui est alors unique, est noté $f'(a)$ ou $\dfrac{df}{dx}(a)$ et appelé la dérivée de f en a.

Dans les écritures des définitions **??** et **??**, on omet de préciser qu'il faut naturellement se limiter à des valeurs de h petites pour lesquelles les points $a+h$ appartiennent au domaine de définition U de f. Ce qui est possible pour un accroissement h petit dans un intervalle ouvert du type $]-r,+r[$ car U étant un ouvert de $\mathbb{R}$ contenant a, il existe un intervalle ouvert du type $]a-r, a+r[$ contenu dans U pour un réel $r > 0$.

Remarque. Cette notion de dérivabilité de f en a est une notion locale, c'est-à-dire ne dépend que de la donnée de f dans un voisinage de a aussi petit soit-il. De plus elle ne dépend pas de la norme choisie sur $\mathbb{R}^p$ puisque les différentes normes sur cet espace sont équivalentes.

On a immédiatement la proposition suivante :

Proposition 53. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Alors la fonction $f : U \subset \mathbb{R} \to \mathbb{R}^p$ est dérivable en a si et seulement si les p composantes $f_j : U \subset \mathbb{R} \to \mathbb{R}$ sont dérivables en a pour $j = 1, \ldots, p$ et dans ce cas*

$$f'(a) = (f_1'(a), \ldots, f_p'(a)).$$

Définition 54. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$. On dit que f est dérivable (ou admet une dérivée) dans une partie V de U si f est dérivable en chaque point de V et la fonction f' définie dans V à valeurs dans $\mathbb{R}^p$ par*

$$f' : V \subset \mathbb{R} \to \mathbb{R}^p : x \mapsto f'(x)$$

est appelée la fonction dérivée de f dans V.

Exemple. La fonction f définie dans $\mathbb{R}$ à valeurs dans $\mathbb{R}^2$ de composantes $f_1 = \cos$ et $f_2 = \sin$, c'est-à-dire

$$f(x) = (\cos x, \sin x) \quad \text{pour} \quad x \in \mathbb{R},$$

est dérivable dans $\mathbb{R}$ de dérivée $f' = (-\sin, \cos)$, c'est-à-dire

$$f'(x) = (-\sin x, \cos x) \quad \text{pour} \quad x \in \mathbb{R}.$$

On définit maintenant la dérivée seconde.

Définition 55. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Si la fonction f admet une dérivée dans un ouvert V contenu dans U et contenant a, et si la fonction dérivée f' est dérivable en a, cette dérivée est notée*

$$f''(a) \quad ou \quad \frac{d^2 f}{dx^2}(a)$$

et appelée la dérivée seconde ou d'ordre deux de f en a.

Par définition on a donc

$$f''(a) = (f')'(a).$$

Extension. Si f est une fonction définie dans un intervalle fermé $[a, b]$ de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$, on définit la notion de dérivée à droite en a (resp. à gauche en b) en considérant de façon classique des limites à droite en a (resp. à gauche en b). On conviendra alors de dire que f est dérivable dans $[a, b]$ si f est dérivable dans $]a, b[$, dérivable à droite en a et dérivable à gauche en b.

1.2. Dérivées partielles. Cas $n \geq 2$

On examine maintenant le cas $\mathbf{n \geq 2}$ et on s'intéresse à des fonctions de n variables réelles $x = (x_1, \ldots, x_n)$.

Si $a = (a_1, \ldots, a_n)$ est un point d'un ouvert U de $\mathbb{R}^n$ et i est un indice compris entre 1 et n, il existe un intervalle ouvert U_i de $\mathbb{R}$ contenant a_i tel que les points

$$(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n)$$

appartiennent à U lorsque le point x_i appartient à U_i (avec des conventions d'écriture évidentes pour $i = 1$ et $i = n$). En effet si par exemple on munit $\mathbb{R}^n$ de la norme

$$\|x\| = \sup_{1 \leq i \leq n} |x_i|$$

il existe $r > 0$ tel que

$$\{x \in \mathbb{R}^n ; \|x - a\| < r\} \subset U$$

et donc on pourra prendre

$$U_i =]a_i - r, a_i + r[.$$

Définition 56. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, a un point de U et i un indice compris entre 1 et n. Si l'application partielle*

$$f(a_1, \ldots, a_{i-1}, ., a_{i+1}, \ldots, a_n) : U_i \subset \mathbb{R} \to \mathbb{R} : x_i \mapsto f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n)$$

est dérivable au point a_i, sa dérivée est alors notée

$$\frac{\partial f}{\partial x_i}(a) \quad ou \quad \partial_i f(a)$$

et appelée la dérivée partielle première ou d'ordre un au point a de f par rapport à la variable x_i.

Comme précédemment pour simplifier dans ce qui suit et quand il n'y aura pas d'ambiguité, on dira *dérivée partielle* à la place de *dérivée partielle première ou d'ordre un.*

Par définition on a donc

$$\frac{\partial f}{\partial x_i}(a) = \frac{dg_i}{dt}(a_i) = g_i'(a_i)$$

si g_i est la fonction définie dans U_i par $g_i(t) = f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$.

Ainsi pour calculer la dérivée partielle par rapport à la variable x_i
- on fixe toutes les autres variables à leur valeur au point considéré,
- puis on dérive la fonction d'une seule variable ainsi obtenue.

D'après la proposition **??**, on a immédiatement la proposition suivante :

Proposition 57. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Alors la fonction $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ a une dérivée partielle en a par rapport à la variable x_i si et seulement si les p composantes $f_j : U \subset \mathbb{R}^n \to \mathbb{R}$ ont des dérivées partielles en a par rapport à la variable x_i pour $j = 1, \ldots, p$ et dans ce cas*

$$\frac{\partial f}{\partial x_i}(a) = \Big(\frac{\partial f_1}{\partial x_i}(a), \ldots, \frac{\partial f_p}{\partial x_i}(a)\Big).$$

On introduit deux notations qui seront utilisées dans la suite.

Définition 58. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et admettant des dérivées partielles en un point a de U.*

Le vecteur $\nabla f(a)$ de $\mathbb{R}^n$ défini par

$$\nabla f(a) = \Big(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\Big)$$

est appelé le gradient de f en a.

Le point a est dit un point critique de f si $\nabla f(a) = 0$.

Définition 59. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et admettant des dérivées partielles en un point a de U. La matrice*

$$Jf(a) = \Big[\frac{\partial f_i}{\partial x_j}(a)\Big]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}}$$

de taille (p, n) est appelée la matrice jacobienne de f en a et si $p = n$ son déterminant est appelé le jacobien de f en a, noté dét$\big(Jf(a)\big)$ ou $\dfrac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}(a)$.

De façon détaillée cette matrice jacobienne s'écrit

$$Jf(a) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \dots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \dfrac{\partial f_2}{\partial x_1}(a) & \dots & \dfrac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_p}{\partial x_1}(a) & \dots & \dfrac{\partial f_p}{\partial x_n}(a) \end{bmatrix}$$

On définit maintenant les dérivées partielles secondes.

Définition 60. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, a un point de U, i et j deux indices compris entre 1 et n.*

Si la fonction f admet une dérivée partielle par rapport à la variable x_j dans un ouvert V contenu dans U et contenant a, et si la fonction dérivée partielle $\dfrac{\partial f}{\partial x_j}$ admet une dérivée partielle en a par rapport à la variable x_i, cette dérivée partielle est alors notée

$$\begin{array}{llll} \dfrac{\partial^2 f}{\partial x_i \partial x_j}(a) & ou \quad \partial^2_{ij} f(a) & si & i \neq j \\ \dfrac{\partial^2 f}{\partial x_i^2}(a) & ou \quad \partial^2_{ii} f(a) & si & i = j \end{array}$$

et appelée la dérivée partielle seconde ou d'ordre deux de f en a par rapport aux variables x_i et x_j.

Par définition on a donc

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_i}(a)$$

ce qui signifie qu'on dérive d'abord par rapport à x_j puis par rapport à x_i.

De façon générale l'ordre des deux dérivations est important. Ainsi par exemple :

Exemple. En notant (x, y) le point générique de $\mathbb{R}^2$ la fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{array}{lll} f(x,y) & = \dfrac{xy(x^2 - y^2)}{x^2 + y^2} & \text{si} \quad (x,y) \neq (0,0) \\ f(x,y) & = 0 & \text{si} \quad (x,y) = (0,0) \end{array}$$

admet des dérivées partielles secondes en $(0, 0)$ avec

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = +1 \quad \text{et} \quad \frac{\partial^2 f}{\partial y \partial x}(0,0) = -1.$$

Cependant sous une hypothèse de continuité, l'ordre dans lequel on dérive est sans importance.

Théorème 61 (de Schwarz). *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$. Si les dérivées partielles $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ et $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ existent dans un ouvert contenant a et sont continues en a, alors*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a).$$

Démonstration. On fait la démonstration dans le cas particulier $n = 2$ et $p = 1$ pour une fonction f définie dans $\mathbb{R}^2$ à valeurs dans $\mathbb{R}$.

On munit $\mathbb{R}^2$ de la norme $\|(x, y)\|_\infty = \sup(|x|, |y|)$.

On suppose que les dérivées partielles secondes $\dfrac{\partial^2 f}{\partial x \partial y}$ et $\dfrac{\partial^2 f}{\partial y \partial x}$ existent dans une boule $\|(x - a, y - b)\|_\infty < r$ pour un $r > 0$ et qu'elles sont continues au point (a, b) de $\mathbb{R}^2$. En particulier les dérivées partielles premières $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$ existent dans cette même boule.

On va montrer que

$$\frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b).$$

Pour h et k fixés tels que $\|(h, k)\|_\infty < r$, on pose

$$\Delta(h, k) = \big(f(a + h, b + k) - f(a + h, b)\big) - \big(f(a, b + k) - f(a, b)\big).$$

La fonction φ définie dans $\mathbb{R}$ à valeurs dans $\mathbb{R}$ par

$$\varphi(x) = f(x, b + k) - f(x, b)$$

admet une dérivée dans l'intervalle $]a - r, a + r[$ donnée par

$$\varphi'(x) = \frac{\partial f}{\partial x}(x, b + k) - \frac{\partial f}{\partial x}(x, b).$$

Donc par la formule des accroissements finis classique rappelée dans la proposition **??**, et appliquée dans l'intervalle $[a, a + h]$, il existe $\theta_1 \in]0, 1[$ tel que

$$\varphi(a + h) - \varphi(a) = h\, \varphi'(a + \theta_1 h)$$

et par suite

$$\Delta(h, k) = h\, \Big(\frac{\partial f}{\partial x}(a + \theta_1 h, b + k) - \frac{\partial f}{\partial x}(a + \theta_1 h, b)\Big).$$

Ensuite la fonction ψ définie dans l'intervalle ouvert $]b - r, b + r[$ par

$$\psi(y) = \frac{\partial f}{\partial x}(a + \theta_1 h, y)$$

admet une dérivée donnée par
$$\psi'(y) = \frac{\partial^2 f}{\partial y \partial x}(a + \theta_1 h, y).$$
Donc par la formule des accroissements finis appliqué dans l'intervalle $[b, b + k]$, il existe un $\theta_2 \in]0, 1[$ tel que
$$\psi(b + k) - \psi(b) = k\, \psi'(b + \theta_2 k)$$
et par suite
$$\Delta(h, k) = hk\, \frac{\partial^2 f}{\partial y \partial x}(a + \theta_1 h, b + \theta_2 k).$$

En refaisant le même raisonnement en échangeant les rôles de x et de y, on montre qu'il existe η_1 et η_2 dans $]0, 1[$ tels que
$$\Delta(h, k) = kh\, \frac{\partial^2 f}{\partial x \partial y}(a + \eta_1 h, b + \eta_2 k).$$

Ainsi pour h et k fixés non nuls dans $\mathbb{R}$ tels que $\|(h, k)\|_\infty < r$, il existe θ_1, θ_2, η_1 et η_2 compris entre 0 et 1 tels que
$$\frac{\partial^2 f}{\partial y \partial x}(a + \theta_1 h, b + \theta_2 k) = \frac{\partial^2 f}{\partial x \partial y}(a + \eta_1 h, b + \eta_2 k).$$

Or comme les fonctions $\dfrac{\partial^2 f}{\partial x \partial y}$ et $\dfrac{\partial^2 f}{\partial y \partial x}$ sont continues en (a, b), pour tout $\varepsilon > 0$ il existe un $\delta > 0$ tel que pour $\|(h', k')\|_\infty < \inf(r, \delta)$ on ait
$$\Big|\frac{\partial^2 f}{\partial x \partial y}(a + h', b + k') - \frac{\partial^2 f}{\partial x \partial y}(a, b)\Big| \leq \varepsilon$$
$$\Big|\frac{\partial^2 f}{\partial y \partial x}(a + h', b + k') - \frac{\partial^2 f}{\partial y \partial x}(a, b)\Big| \leq \varepsilon,$$
et donc si $\|(h, k))\|_\infty < \inf(r, \delta)$ on a
$$\Big|\frac{\partial^2 f}{\partial x \partial y}(a + \eta_1 h, b + \eta_2 k) - \frac{\partial^2 f}{\partial x \partial y}(a, b)\Big| \leq \varepsilon$$
$$\Big|\frac{\partial^2 f}{\partial y \partial x}(a + \theta_1 h, b + \theta_2 k) - \frac{\partial^2 f}{\partial y \partial x}(a, b)\Big| \leq \varepsilon.$$
Par suite pour tout $\varepsilon > 0$ on a
$$\Big|\frac{\partial^2 f}{\partial y \partial x}(a, b) - \frac{\partial^2 f}{\partial x \partial y}(a, b)\Big| \leq 2\,\varepsilon,$$
autrement dit
$$\frac{\partial^2 f}{\partial y \partial x}(a, b) = \frac{\partial^2 f}{\partial x \partial y}(a, b). \quad \diamond$$

On introduit enfin la notation suivante :

Définition 62. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et admettant des dérivées partielles d'ordre 2 en un point a de U. La matrice*

$$Hf(a) = \Big[\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\Big]_{\substack{i=1,\dots,n\\ j=1,\dots,n}}$$

de taille (n, n) est appelée la matrice hessienne de f en a.

De façon détaillée cette matrice hessienne s'écrit

$$Hf(a) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(a) & \dots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \dots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(a) & \dots & \dfrac{\partial^2 f}{\partial^2 x_n}(a) \end{bmatrix}$$

Convention. Les formules sont et seront écrites de façon générale pour $n \geq 1$ en convenant que si $n = 1$ les notions de dérivées partielles et de dérivées coïncident, c'est-à-dire

$$\frac{\partial f}{\partial x_1} = \frac{df}{dx} = f' \quad \text{et} \quad \frac{\partial^2 f}{\partial x_1^2} = \frac{d^2 f}{dx^2} = f''.$$

2. Différentielles d'ordre un

On s'intéresse maintenant dans ce paragraphe à une notion de dérivation « de type global » par rapport aux n variables réelles pour des fonctions de n variables réelles pour $n \geq 1$.

2.1. Définition

La définition **??** ne peut pas être étendue à une fonction définie dans un ouvert U de $\mathbb{R}^n$ pour $n \geq 2$ car on ne peut pas diviser par un élément h de $\mathbb{R}^n$. Par contre la définition **??** (qui lui est équivalente) peut être étendue sous la forme suivante :

Définition 63. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. On dit que f est une fois différentiable en a s'il existe une application linéaire L de $\mathbb{R}^n$ dans $\mathbb{R}^p$ telle que pour une norme $\|\cdot\|$ sur $\mathbb{R}^n$*

$$f(a + h) = f(a) + L(h) + \|h\|\,\varepsilon(h)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, continue en 0 et nulle en 0. Cette application linéaire L qui est alors unique, est notée $df(a)$ et appelée la différentielle première ou d'ordre un de f en a.

Pour simplifier on dira *différentiable* à la place de *une fois différentiable* et *différentielle* à la place de *différentielle première ou d'ordre un.*

Comme dans la définition **??** cette formule a bien un sens pour un accroissement h petit dans un voisinage de 0 dans $\mathbb{R}^n$.

Démonstration. (de l'unicité) Soit L_1 et L_2 deux applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^p$ telles que pour h petit dans $\mathbb{R}^n$

$$f(a+h) = f(a) + L_1(h) + \|h\|\,\varepsilon_1(h) \quad \text{et} \quad f(a+h) = f(a) + L_2(h) + \|h\|\,\varepsilon_2(h).$$

Pour $h \neq 0$ dans $\mathbb{R}^n$ fixé, on considère un accroissement de la forme th pour t réel petit. D'après les formules précédentes on a donc

$$L_1(th) - L_2(th) = \|th\|\,\varepsilon_2(th) - \|th\|\,\varepsilon_1(th),$$

soit pour $t \neq 0$

$$L_1(h) - L_2(h) = \frac{\|th\|\,\varepsilon_2(th) - \|th\|\,\varepsilon_1(th))}{t}.$$

Or pour une norme $\|\cdot\|_p$ de $\mathbb{R}^p$ on a

$$\left\| \frac{\|th\|\,\varepsilon_2(th) - \|th\|\,\varepsilon_1(th)}{t} \right\|_p \leq \|h\|\,(\|\varepsilon_2(th)\|_p + \|\varepsilon_1(th)\|_p).$$

En faisant tendre t vers 0 dans $\mathbb{R}$ on en déduit que $L_1(h) - L_2(h) = 0$. Ainsi $L_1 = L_2$. $\diamond$

Remarque. Comme pour la notion de dérivabilité d'une fonction d'une variable réelle, cette notion de différentiabilité d'une fonction de n variables réelles est une notion locale et ne dépend pas des normes données sur $\mathbb{R}^n$ et sur $\mathbb{R}^p$.

Comme pour la dérivabilité des fonctions d'une variable réelle on a

Proposition 64. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Alors la fonction $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ est différentiable en a si et seulement si les p composantes $f_j : U \subset \mathbb{R}^n \to \mathbb{R}$ sont différentiables en a pour $j = 1, \cdots, p$ et dans ce cas $df(a)$ a pour composantes $df_1(a), \ldots, df_p(a)$.*

Ce que l'on note

$$df(a) = (df_1(a), \ldots, df_p(a))$$

dans le sens que pour h dans $\mathbb{R}^n$, $df(a) \cdot h$ est le vecteur de $\mathbb{R}^n$ défini par

$$df(a)h = (df_1(a)h, \ldots, df_p(a)h).$$

Définition 65. *Une fonction f définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ est dite différentiable (ou admet une différentielle) dans une partie V de U si elle est différentiable en tout point de V et on note df l'application de V dans l'espace $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^p)$ définie par*

$$df : V \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^p) : x \mapsto df(x).$$

Exemple. (cas constant) Soit f une application constante de $\mathbb{R}^n$ dans $\mathbb{R}^p$, c'est-à-dire il existe un $\lambda \in \mathbb{R}^p$ tel que $f(x) = \lambda$ pour tout $x \in \mathbb{R}^n$. Cette fonction f pourra être notée simplement λ.

Pour tous a et h dans $\mathbb{R}^n$ on a donc

$$f(a+h) = f(a)$$

et par suite on peut écrire

$$f(a+h) = f(a) + Lh + \|h\| \varepsilon(h)$$

où L est l'application linéaire constante nulle définie par

$$Lh = 0$$

et $\varepsilon(h) = 0$, où 0 désigne ici l'élément neutre de l'addition dans $\mathbb{R}^p$.

On en déduit donc que f est différentiable en a et, en notant simplement 0 cette application L, que $df(a) = 0$.

Exemple. (cas linéaire) Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Pour tous a et h dans $\mathbb{R}^n$ on a donc

$$f(a+h) = f(a) + f(h)$$

et par suite

$$f(a+h) = f(a) + Lh + \|h\| \varepsilon(h)$$

où L est l'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ définie par

$$Lh = f(h)$$

et comme précédemment $\varepsilon(h) = 0$.

On en déduit donc que f est différentiable en a avec $df(a) = f$.

On peut écrire ce résultat sous forme matricielle. Relativement à la base canonique de $\mathbb{R}^n$, cette application f est représentée par une matrice A de taille (p, n) sous la forme

$$f(x) = Ax$$

pour $x \in \mathbb{R}^n$.

D'après ce qui précède $df(a)$ est représentée par la matrice A, ce qu'on s'écrit matriciellement

$$df(a)h = Ah$$

pour $h \in \mathbb{R}^n$.

Exemple. (cas quadratique) Soit f une forme quadratique sur $\mathbb{R}^n$. Relativement à la base canonique de $\mathbb{R}^n$, cette application f est représentée par une matrice symétrique A de taille (n, n) sous la forme

$$f(x) = < Ax, x >$$

pour $x \in \mathbb{R}^n$.

Pour tous a et h dans $\mathbb{R}^n$ on a donc

$$f(a+h) = f(a) + 2 < Aa, h > + < Ah, h >$$

et par suite

$$f(a+h) = f(a) + Lh + \|h\| \varepsilon(h)$$

où L est la forme linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$Lh = 2 < Aa, h >$$

et $\|h\| \varepsilon(h) =< Ah, h >$. En effet le terme $< Ah, h >$ est bien un $\|h\| \varepsilon(h)$ puisque par le corollaire **??** du chapitre 2, il existe une constante C telle que $| < Ah, h > | \leq C \|h\|^2$.

On en déduit donc que f est différentiable en a avec

$$df(a)h = 2 < Aa, h > .$$

Autrement dit la forme linéaire $df(a)$ de $\mathbb{R}^n$ dans $\mathbb{R}$ est représentée relativement aux bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}$ par la matrice transposée de $2Aa$, de taille $(1, n)$.

2.2. Liens entre dérivabilité et différentiabilité. Condition nécessaire et condition suffisante de différentiabilité

Tout d'abord pour $n \geq 1$ on a la condition nécessaire suivante :

Théorème 66. *(Condition nécessaire) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Si f est différentiable en a alors* **nécessairement** *f admet des dérivées partielles en a données par*

$$\frac{\partial f}{\partial x_i}(a) = df(a)e_i$$

pour $i = 1, \ldots, n$, soit

$$\frac{\partial f_j}{\partial x_i}(a) = \big(df(a)e_i\big)_j$$

pour $i = 1, \ldots, n$ et $j = 1, \ldots, p$.

Ainsi dans ce cas, relativement aux bases canoniques de $\mathbb{R}^n$ et de $\mathbb{R}^p$, l'application linéaire $df(a)$ de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est représentée par la matrice jacobienne $Jf(a)$.

Autrement dit la différentielle $df(a)$ est l'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ qui s'écrit sous la forme maticielle

$$df(a)h = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \ldots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \dfrac{\partial f_2}{\partial x_1}(a) & \ldots & \dfrac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_p}{\partial x_1}(a) & \ldots & \dfrac{\partial f_p}{\partial x_n}(a) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ \vdots \\ h_n \end{bmatrix}$$

pour $h = (h_1, \ldots, h_n)$ appartenant à $\mathbb{R}^n$.

Démonstration. On note $\|\cdot\|_n$ et $\|\cdot\|_p$ des normes sur les espaces $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement.

Pour t réel assez petit pour que $a+te_i$ appartienne à U on a par la définition de la différentielle de f en a

$$\begin{aligned} f(a_1,\ldots,a_{i-1},a_i+t,a_{i+1}\ldots,a_n) &= f(a+te_i) = f(a)+df(a)\,(te_i)+\|te_i\|_n\,\varepsilon(te_i) \\ &= f(a_1,\ldots,a_{i-1},a_i,a_{i+1}\ldots,a_n)+t\,df(a)\,e_i+\|te_i\|_n\,\varepsilon(te_i). \end{aligned}$$

Or

$$\frac{\|\,\|te_i\|_n\,\varepsilon(te_i)\,\|_p}{|t|} = \|e_i\|_n\,\|\varepsilon(te_i)\|_p$$

qui tend vers 0 quand t tend vers 0 dans $\mathbb{R}$, $t\neq 0$.

Par la définition **??** on en déduit que l'application partielle

$$f(a_1,\ldots,a_{i-1},.,a_{i+1},\ldots,a_n) : U_i\subset\mathbb{R}\to\mathbb{R}^p : x_i\mapsto f(a_1,\ldots,a_{i-1},x_i,a_{i+1},\ldots,a_n)$$

est dérivable au point a_i, de dérivée $df(a)e_i$, c'est-à-dire que

$$\frac{\partial f}{\partial x_i}(a) = df(a)e_i.$$

Ainsi

$$\big(df(a)e_i\big)_j = \frac{\partial f_j}{\partial x_i}(a)$$

pour $i=1,\ldots,n$ et $j=1,\ldots,p$, d'où l'on déduit la représentation matricielle de l'application linéaire $df(a)$ de $\mathbb{R}^n$ dans $\mathbb{R}^p$. $\diamond$

On donne deux cas particuliers.

- Pour $n=1$, si f est une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et différentiable en un point a de U, la différentielle $df(a)$ est l'application linéaire de $\mathbb{R}$ dans $\mathbb{R}^p$ qu'on écrit sous la forme matricielle

$$df(a)h = \begin{bmatrix} f'_1 \\ \vdots \\ f'_n \end{bmatrix} \begin{bmatrix} h \end{bmatrix}$$

pour tout réel h de $\mathbb{R}$, c'est-à-dire aussi comme égalité entre vecteurs de $\mathbb{R}^p$

$$df(a)h = h\,f'(a)$$

qui est la multiplication du vecteur $f'(a)$ de $\mathbb{R}^p$ par le scalaire h.

- Pour $p=1$, si f est une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et différentiable en un point a de U, la différentielle $df(a)$ est l'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$ qu'on écrit sous la forme matricielle

$$df(a)h = \Big[\frac{\partial f}{\partial x_1}(a)\cdots\frac{\partial f}{\partial x_n}(a)\Big] \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}$$

pour tout vecteur h de $\mathbb{R}^n$ de composantes $h_1, \dots, h_n$, c'est-à-dire aussi comme égalité entre nombres de $\mathbb{R}$

$$df(a)h =< \nabla f(a), h >= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\, h_i.$$

Notation différentielle. L'application p_i de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$p_i(x) = x_i$$

pour $x = (x_1, \dots, x_n)$ dans $\mathbb{R}^n$, est linéaire donc différentiable dans $\mathbb{R}^n$ avec

$$dp_i(a) = p_i$$

pour tout $a \in \mathbb{R}^n$. Autrement dit l'application différentielle dp_i est constante en tant qu'application de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n; \mathbb{R})$ égale à p_i.

Comme $p_i(x) = x_i$, il est habituel de noter la différentielle $dp_i(a)$ sous la forme dx_i.

Ainsi si par exemple f est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ différentiable en a alors

$$df(a)(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\, h_i = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\, dx_i(h) = (\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\, dx_i)(h)$$

d'où l'égalité entre applications linéaires

$$df(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\, dx_i$$

où rappelons que dx_i est l'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$dx_i(h) = h_i$$

pour $h = (h_1, \dots, h_n)$.

Inversement pour $n = 1$ la condition nécessaire du théorème **??** est en fait suffisante. Ainsi pour une fonction d'une variable réelle, on a l'équivalence suivante :

Théorème 67. *(n=1. Condition nécessaire et suffisante) Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}^p$ et a un point de U. La fonction f est différentiable en a si et seulement si elle est dérivable en a et dans ce cas pour tout $h \in \mathbb{R}$ on a*

$$df(a)(h) = h\, f'(a).$$

Par contre si $n \geq 2$, l'existence seule de toutes les dérivées partielles de f en a est une condition nécessaire mais non suffisante en général pour que f soit différentiable a.

Exemple. La fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{array}{lll} f(x,y) = & \dfrac{xy}{x^2+y^2} & \text{si} \quad (x,y) \neq (0,0) \\ f(x,y) = & 0 & \text{si} \quad (x,y) = (0,0) \end{array}$$

admet des dérivées partielles par rapport à x et à y en $(0,0)$ et cependant n'est pas différentiable en $(0,0)$. En effet comme ces dérivées partielles sont nulles, si f était différentiable en $(0,0)$, on pourrait alors nécessairement écrire pour tout $(x,y) \neq (0,0)$

$$\frac{xy}{x^2+y^2} = \|(x,y)\|\,\varepsilon(x,y)$$

donc en particulier pour tout $x \neq 0$ (et $y = x$) on aurait

$$\frac{1}{2} = \|(x,x)\|\,\varepsilon(x,x),$$

ce qui est impossible (en faisant tendre x vers 0).

Exemple. La fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{array}{ll} f(x,y) = & 1 \quad \text{si} \quad x=0 \quad \text{ou} \quad y=0 \\ f(x,y) = & 0 \quad \text{sinon} \end{array}$$

admet des dérivées partielles par rapport à x et à y en $(0,0)$ et cependant n'est pas différentiable en $(0,0)$.

Pour $n \geq 2$, on donne maintenant une condition suffisante de différentiabilité portant sur la continuité des dérivées partielles, qui sera en fait **le critère pratique** le plus commode de différentiabilité.

Théorème 68. *($n \geq 2$. Condition suffisante) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Si toutes les dérivées partielles de f existent dans un voisinage de a et sont continues en a alors f est différentiable en a.*

Démonstration. On fait la démonstration dans le cas particulier où $n = 2$ pour une fonction f définie dans $\mathbb{R}^2$ à valeurs dans $\mathbb{R}$.

On munit $\mathbb{R}^2$ de la norme

$$\|(x,y)\|_\infty = \sup(|x|,|y|).$$

On suppose que les dérivées partielles $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$ de f existent dans la boule ouverte $\|(x-a,y-b)\|_\infty < r$ pour un $r > 0$ et sont continues au point (a,b) de $\mathbb{R}^2$. On va montrer que f est différentiable en (a,b).

Pour (h,k) fixé tel que $\|(h,k)\|_\infty < r$, on décompose

$$f(a+h,b+k) = f(a,b) + \big(f(a+h,b+k) - f(a+h,b)\big) + \big(f(a+h,b) - f(a,b)\big).$$

La fonction φ_1 définie dans $\mathbb{R}$ à valeurs dans $\mathbb{R}$ par

$$\varphi_1(y) = f(a+h,y)$$

admet une dérivée dans l'intervalle ouvert $]b-r,b+r[$ donnée par

$$\varphi_1'(y) = \frac{\partial f}{\partial y}(a+h,y).$$

Donc par la formule des accroissements finis dans l'intervalle fermé d'extrémités b et $b+k$ (contenu dans $]b-r, b+r[$) il existe $\theta_1 \in]0,1[$ tel que

$$\varphi_1(b+k) - \varphi_1(b) = k\,\varphi_1'(b+\theta_1 k)$$

soit

$$f(a+h, b+k) - f(a+h, b) = k\,\frac{\partial f}{\partial y}(a+h, b+\theta_1 k).$$

Or comme la dérivée partielle $\dfrac{\partial f}{\partial y}$ est continue au point (a,b), pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour $\|(h', k')\|_\infty < \inf(r, \delta)$ on ait

$$|\frac{\partial f}{\partial y}(a+h', b+k') - \frac{\partial f}{\partial y}(a,b)| \leq \varepsilon$$

donc pour $\|(h,k)\|_\infty < \inf(r,\delta)$

$$|k\frac{\partial f}{\partial y}(a+h, b+\theta_1 k) - k\frac{\partial f}{\partial y}(a,b)| \leq \varepsilon|k| \leq \varepsilon\|(h,k)\|_\infty.$$

Ainsi on peut écrire

$$f(a+h, b+k) - f(a+h, b) = k\,\frac{\partial f}{\partial y}(a,b) + \|(h,k)\|_\infty\,\varepsilon_1((h,k)).$$

En procédant de même avec la variable x en introduisant la fonction

$$\varphi_2(x) = f(x, b)$$

on peut écrire

$$f(a+h, b) - f(a,b) = h\,\frac{\partial f}{\partial x}(a,b) + \|(h,k)\|_\infty\,\varepsilon_2((h,k)).$$

En regroupant ces deux termes on en déduit que

$$f(a+h, b+k) = f(a,b) + h\,\frac{\partial f}{\partial x}(a,b) + k\,\frac{\partial f}{\partial y}(a,b) + \|(h,k)\|_\infty\,\varepsilon((h,k)).$$

Ainsi f est différentiable en (a,b) et on retrouve (nécessairement) que

$$df(a,b)\,(h,k) = h\,\frac{\partial f}{\partial x}(a,b) + k\,\frac{\partial f}{\partial y}(a,b). \quad \diamond$$

Ceci nous amène à poser la définition suivante :

Définition 69. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$.*

On dit que f est continûment différentiable dans une partie V de U ou de classe C^1 dans V si toutes les dérivées partielles de f existent dans un ouvert contenant V et sont continues dans V.

2.3. Propriétés

Proposition 70. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et a un point de U. Si f est différentiable en a alors f est continue en a.*

Démonstration. Pour des normes $\|\cdot\|_n$ et $\|\cdot\|_p$ sur $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement on a par la différentiabilité de f en a

$$f(a+h) = f(a) + df(a)(h) + \|h\|_n \varepsilon(h)$$

avec

$$\|df(a)(h)\|_p \leq C\|h\|_n \quad \text{et} \quad \|\varepsilon(h)\|_p \leq C$$

pour une constante C. D'où l'on déduit la continuité de f en a. $\diamond$

Exemple. On considère la fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{aligned} f(x,y) = &\ \frac{xy}{x^2+y^2} \quad \text{si} \quad (x,y) \neq (0,0) \\ f(x,y) = &\ 0 \qquad\quad\ \text{si} \quad (x,y) = (0,0). \end{aligned}$$

Cette fonction f n'est pas différentiable en $(0,0)$ puisqu'elle n'est pas continue en $(0,0)$.

Proposition 71. *1- Soit f et g deux fonctions définies dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$, k une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U.*

1 - Si f et g sont différentiables en a, alors la fonction $f+g$ est différentiable en a avec

$$d(f+g)(a) = df(a) + dg(a).$$

2 - Si f et k sont différentiables en a, alors la fonction kf est différentiable en a avec

$$d(kf)(a) = k(a)\ df(a) + dk(a)(\cdot)\, f(a).$$

3 - Si k est non nulle en a, alors il existe un ouvert V contenu dans U et contenant a tel que $k(x) \neq 0$ pour $x \in V$ et la fonction $\frac{1}{k}$ est différentiable en a avec

$$d\Big(\frac{1}{k}\Big)(a) = -\frac{1}{k^2(a)}\, dk(a).$$

Démonstration. On démontre par exemple la propriété pour le produit. Par la différentiabilité de f et k en a on a

$$f(a+h) = f(a) + df(a)(h) + \|h\|_n \varepsilon_f(h) \quad \text{et} \quad k(a+h) = k(a) + dk(a)(h) + \|h\|_n \varepsilon_k(h)$$

où ε_f et ε_k sont des fonctions définies dans un voisinage de 0 dans $\mathbb{R}^n$, continues et nulles en 0 ; de plus

$$\|df(a)(h)\|_p \leq C\|h\|_n \quad \text{et} \quad |dk(a)(h)| \leq C\|h\|_n$$

pour une constante C. Par suite

$$k(a+h)\ f(a+h) = k(a)\ f(a) + k(a)\, df(a)(h) + dk(a)(h)\, f(a) + \|h\|_n r(h)$$

avec

$$r(h) = k(a)\,\varepsilon_f(h) + \varepsilon_k(h)\,f(a) + dk(a)(h)\,df(a)(h) + dk(a)\,\varepsilon_f(h) + \varepsilon_k(h)\,df(a)(h) + \|h\|_n\,\varepsilon_k(h)\,\varepsilon_f(h).$$

Ce terme $r(h)$ est donc du type $\varepsilon(h)$ d'après les propriétés précédentes, et comme de plus l'application $k(a)\,df(a) + dk(a)(\cdot)\,f(a)$ est linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, on déduit la différentiabilité de kf en a et la formule de sa différentielle :

$$d(kf)(a)h = k(a)\ df(a)h + dk(a)h\,f(a).$$

$\diamond$

Proposition 72. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et g une fonction définie dans un ouvert V de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^q$ tel que $f(U) \subset V$. Si f est différentiable en un point a de U et si g est différentiable en $f(a)$ alors la fonction $g \circ f$ est différentiable en a avec*

$$d(g \circ f)(a) = dg(f(a)) \circ df(a).$$

Démonstration. Par la différentiabilité de f en a et de g en $f(a)$ on a

$$f(a+h) = f(a) + df(a)(h) + \|h\|_n\,\varepsilon_f(h)$$

$$\begin{aligned} g(f(a+h)) \quad = \quad & g(f(a)) + dg(f(a))\big(df(a)(h) + \|h\|_n\,\varepsilon_f(h)\big) \\ & +\big(\|df(a)(h) + \|h\|_n\,\varepsilon_f(h)\|_p\big)\,\|\,\varepsilon_g(df(a)(h) + \|h\|_n\,\varepsilon_f(h)) \end{aligned}$$

où ε_f et ε_g sont des fonctions définies dans un voisinage de 0 dans $\mathbb{R}^n$ et $\mathbb{R}^p$ respectivement, continues et nulles en 0, et

$$\|df_a(h)\|_p \leq C\|h\|_n \quad \text{et} \quad \|dg(f(a))(k)\|_q \leq C\|k\|_p$$

pour une constante C.

Par suite on peut écrire

$$(g \circ f)(a+h) = (g \circ f)(a) + dg(f(a))\big(df(a)\,h\big) + \|h\|_n\,\varepsilon(h)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue et nulle en 0. Comme de plus l'application

$$dg(f(a)) \circ df(a) : h \to dg(f(a))\big(df(a)(h)\big)$$

est linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^q$, on en déduit la différentiabilité de $g \circ f$ en a et la formule de sa différentielle. $\diamond$

Concernant les dérivées partielles des fonctions composées, on déduit de la proposition précédente :

Corollaire 73. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et g une fonction définie dans un ouvert V de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^q$ tel que $f(U) \subset V$. Si f est différentiable en un point de a de U et si g est différentiable en $f(a)$, alors*

1 - chaque composante f_j de f admet des dérivées partielles $\dfrac{\partial f_j}{\partial x_i}(a)$ en a par rapport à chaque variable x_i pour $j = 1, \dots, p$ et $i = 1, \dots, n$,

2 - chaque composante g_k de g admet des dérivées partielles $\dfrac{\partial g_k}{\partial y_j}(f(a))$ en $f(a)$ par rapport à chaque variable y_j pour $k = 1, \dots, q$ et $j = 1, \dots, p$,

3 - chaque composante $(g \circ f)_k$ de la fonction composée $g \circ f$ admet des dérivées partielles $\dfrac{\partial (g\circ f)_k}{\partial x_i}(a)$ en a par rapport à chaque variable x_i pour $k = 1, \dots, q$ et $i = 1, \dots, n$, données par les formules

$$\frac{\partial (g\circ f)_k}{\partial x_i}(a) = \sum_{j=1}^{p} \frac{\partial g_k}{\partial y_j}(f(a))\, \frac{\partial f_j}{\partial x_i}(a).$$

Démonstration. D'après la proposition **??** on a

$$d(g \circ f)(a) = dg(f(a)) \circ df(a).$$

Chacune de ces applications linéaires est représentée relativement aux bases canoniques de $\mathbb{R}^n$, $\mathbb{R}^p$ et $\mathbb{R}^q$ par une matrice jacobienne et la matrice $J(g \circ f)(a)$ associée à la composée $dg(f(a)) \circ df(a)$ est donc le produit des matrices $Jg(f(a))$ et $Jf(a)$ associées respectivement à $dg(f(a))$ et $df(a)$. Ainsi

$$J(g \circ f)(a) = Jg(f(a)) \times Jf(a)$$

c'est-à-dire

$$\left[\frac{\partial (g\circ f)_k}{\partial x_i}(a)\right]_{\substack{k=1,\dots,q\\ i=1,\dots,n}} = \left[\frac{\partial g_k}{\partial y_j}(f(a))\right]_{\substack{k=1,\dots,q\\ j=1,\dots,p}} \times \left[\frac{\partial f_j}{\partial x_i}(a)\right]_{\substack{j=1,\dots,p\\ i=1,\dots,n}}$$

d'où l'on déduit la forme annoncée des cœfficients de la matrice jacobienne $J(g \circ f)(a)$. $\diamond$

3. Différentielle d'ordre deux

On s'intéresse maintenant aux fonctions deux fois différentiables en se limitant aux fonctions à valeurs dans $\mathbb{R}$. Pour les fonctions à valeurs dans $\mathbb{R}^p$, on raisonne sur chaque composante.

3.1. Définitions

Pour les fonctions d'une variable réelle, on a défini la notion de dérivée d'ordre deux par itération de la notion de dérivée d'ordre un. On est donc tenté maintenant de définir également la notion de différentielle d'ordre deux par itération de la notion de différentielle d'ordre un de la façon suivante :

Définition formelle. *Avec les notations habituelles, on dit que f est deux fois différentiable en a si*

1 - la fonction f est différentiable dans un ouvert V contenu dans U et contenant a,

2 - la fonction df définie dans V à valeurs dans $\mathcal{L}(\mathbb{R}^n;\mathbb{R})$ est différentiable en a.

La différentielle $d(df)(a)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n;\mathbb{R})$, notée $d^2f(a)$ et appelée la différentielle d'ordre deux de f en a.

Ainsi avec cette définition l'application

$$h \to d^2f(a)(h)$$

est linéaire de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n;\mathbb{R})$ et pour tout $h \in \mathbb{R}^n$, l'application

$$k \to \big(d^2f(a)(h)\big)(k)$$

est linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$.

Par conséquent l'application

$$(h,k) \to \big(d^2f(a)(h)\big)(k)$$

est une forme bilinéaire sur $\mathbb{R}^n \times \mathbb{R}^n$. On peut donc identifier cette différentielle d'ordre deux $d^2f(a)$ à une forme bilinéaire sur $\mathbb{R}^n \times \mathbb{R}^n$ que l'on notera encore $d^2f(a)$ définie par

$$d^2f(a)(h,k) = \big(d^2f(a)(h)\big)(k)$$

pour tout $(h,k) \in \mathbb{R}^n \times \mathbb{R}^n$.

Cependant cette présentation est **formelle** car par la définition **??** on n'a défini une notion de différentielle (d'ordre un) que pour des fonctions à valeurs dans un espace du type $\mathbb{R}^p$ et non pas pour des fonctions à valeurs dans un espace du type $\mathcal{L}(\mathbb{R}^n;\mathbb{R})$ comme on l'a fait ici pour la fonction différentielle df.

Pour pouvoir utiliser la définition **??**, on va faire intervenir la représentation matricielle de la différentielle df relativement à la base canonique de $\mathbb{R}^n$ par le gradient ∇f et ainsi se ramener à une fonction à valeurs dans $\mathbb{R}^n$.

Par le théorème **??** on rappelle que si la fonction f est différentiable dans un ouvert V de $\mathbb{R}^n$, elle admet des dérivées partielles $\dfrac{\partial f}{\partial x_j}$ dans V pour $j = 1,\dots,n$ et on peut définir la fonction ∇f de composantes $\dfrac{\partial f}{\partial x_j}$ comme fonction définie dans V à valeurs dans $\mathbb{R}^n$, soit

$$\nabla f = \Big(\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}\Big).$$

On peut alors poser

Définition 74. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et a un point de U. On dit que f est deux fois différentiable en a si*

1 - la fonction f est différentiable dans un ouvert V de $\mathbb{R}^n$ contenu dans U et contenant a,

2 - la fonction ∇f définie dans l'ouvert V de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^n$ est différentiable en a.

La différentielle d'ordre deux ou seconde de f en a est la forme bilinéaire sur $\mathbb{R}^n \times \mathbb{R}^n$ notée $d^2f(a)$ et définie pour $(h,k) \in \mathbb{R}^n \times \mathbb{R}^n$ par

$$d^2f(a)(h,k) = < d(\nabla f)(a)(h), k > .$$

3.2. Lien entre dérivabilité et différentiabilité à l'ordre deux. Condition nécessaire et condition suffisante de différentiabilité

On a d'abord la condition nécessaire suivante pour $n \geq 1$ (avec la convention pour $n = 1$) :

Théorème 75. *($n \geq 1$. Condition nécessaire) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et a un point de U. Si f est deux fois différentiable en a, alors* **nécessairement** *la fonction f admet des dérivées partielles secondes en a données par*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = d^2f(a)(e_j, e_i)$$

pour i, $j = 1, \ldots, n$.

Ainsi dans ce cas relativement à la base canonique de $\mathbb{R}^n$, la forme bilinéaire $d^2f(a)$ sur $\mathbb{R}^n \times \mathbb{R}^n$ est représentée par la matrice hessienne $Hf(a)$.

Autrement dit sous forme détaillée

$$d^2f(a)(h,k) = < Hf(a)\,h, k > = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_j k_i$$

pour h et $k \in \mathbb{R}^n$.

Démonstration. La matrice jacobienne $J(\nabla f)(a)$ de l'application ∇f est la matrice hessienne $Hf(a)$ et donc par le théorème **??** on peut écrire sous forme matricielle

$$d(\nabla f)(a)(h) = Hf(a)\,h$$

pour $h \in \mathbb{R}^n$. ◇

Exemple. (cas quadratique) Soit f une forme quadratique sur $\mathbb{R}^n$.

Relativement à la base canonique de $\mathbb{R}^n$, cette application f est représentée par une matrice symétrique A de taille (n, n) sous la forme

$$f(x) = < Ax, x >$$

pour $x \in \mathbb{R}^n$.

On a montré que f est différentiable dans $\mathbb{R}^n$ et que l'application linéaire $df(x)$ de $\mathbb{R}^n$ dans $\mathbb{R}$ est représentée relativement aux bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}$ par la matrice jacobienne $Jf(x) = 2Ax$ de taille $(n, 1)$.

Ainsi l'application gradient de f est donc donnée sous forme matricielle par

$$\nabla f(x) = 2Ax.$$

Cette application ∇f est linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$, donc différentiable dans $\mathbb{R}^n$ de différentielle ∇f, c'est-à-dire

$$d(\nabla f)(x)(h) = \nabla f(h) \quad \text{pour} \quad x,\, h \in \mathbb{R}^n$$

Par suite f est deux fois différentiable dans $\mathbb{R}^n$ avec

$$d^2 f(x)(h,k) =< \nabla f(h), k >= 2 < Ah, k > \quad \text{pour} \quad x,\, h, k \in \mathbb{R}^n.$$

En particulier

$$Hf(x) = 2A \quad \text{pour} \quad x \in \mathbb{R}^n.$$

Inversement pour une fonction d'une variable réelle ($n = 1$), la condition nécessaire du théorème **??** est suffisante et on a l'équivalence suivante :

Proposition 76. *($n = 1$. Condition nécessaire et suffisante) Soit f une fonction définie dans un ouvert U de $\mathbb{R}$ à valeurs dans $\mathbb{R}$ et a un point de U. La fonction f est deux fois différentiable en a* **si et seulement si** *elle est deux fois dérivable en a et dans ce cas pour tous h et $k \in \mathbb{R}$ on a*

$$d^2 f(a)(h,k) = h\,k\,f''(a).$$

Autrement dit la différentielle $d^2 f(a)$ est la forme bilinéaire sur $\mathbb{R} \times \mathbb{R}$ qui s'écrit sous la forme

$$\mathbb{R} \to \mathbb{R} : (h,k) \mapsto h\,k\,f''(a)$$

qui est la multiplication du réel $f''(a)$ de $\mathbb{R}$ par le scalaire hk.

Par contre pour $n \geq 2$, la condition nécessaire du théorème **??** n'est pas suffisante en général. On donne maintenant une condition suffisante portant sur la continuité des dérivées partielles d'ordre deux, qui sera en fait le critère pratique le plus commode de différentiabilité d'ordre 2.

Théorème 77. *($n \geq 2$. Condition suffisante) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et a un point de U. Si toutes les dérivées partielles d'ordre deux de f existent dans un ouvert contenant a et sont continues en a, alors f est deux fois différentiable en a.*

Démonstration. Cette propriété est un corollaire de la définition **??** et du théorème **??**.
$\diamond$

Ceci nous amène à poser la définition suivante :

Définition 78. *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$.*

On dit que f est deux fois continûment différentiable dans une partie V de U, ou de classe C^2 dans V, si toutes les dérivées partielles d'ordre deux de f existent dans un ouvert contenant V et sont continues dans V.

Pour terminer on donne une propriété importante des différentielles secondes portant sur leur symétrie.

Théorème 79 (de Schwarz). *Pour $n \geq 2$, soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et a un point de U. Si la fonction f est deux fois différentiable en a, alors la forme bilinéaire $d^2f(a)$ sur $\mathbb{R}^n \times \mathbb{R}^n$ est symétrique, c'est-à-dire*

$$d^2f(a)\,(h,k) = d^2f(a)\,(k,h) \quad \textit{pour} \quad (h,k) \in \mathbb{R}^n \times \mathbb{R}^n,$$

soit aussi

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a) \quad \textit{pour} \quad i,\, j = 1, \ldots, n.$$

Démonstration. On a montré dans le théorème **??** un résultat plus faible en supposant a priori que f est deux fois continûment différentiable en a. On admettra le résultat dans le cas général. $\diamond$

Exemple. La fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$\begin{aligned} f(x,y) &= \frac{xy(x^2-y^2)}{x^2+y^2} && \text{si} \quad (x,y) \neq (0,0) \\ f(x,y) &= 0 && \text{si} \quad (x,y) = (0,0) \end{aligned}$$

admet des dérivées partielles secondes en $(0,0)$ et cependant n'est pas deux fois différentiable en $(0,0)$ puisque

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = +1 \quad \text{et} \quad \frac{\partial^2 f}{\partial y \partial x}(0,0) = -1.$$

4. Formules des accroissements finis et de Taylor

On rappelle que si une fonction f définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ est une fois différentiable en un point a de U, alors

$$f(a+h) = f(a) + df(a)\,h + \|h\|\,\varepsilon(h)$$

où la différentielle $df(a)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Ainsi la différentielle d'ordre un permet de donner une approximation de type **linéaire** d'une fonction au voisinage d'un point, et de même la différentielle d'ordre deux permet de donner une approximation de type **quadratique**. On va préciser ces approximations d'ordre un et d'ordre deux dans certains cas.

Si a et b sont deux points de $\mathbb{R}^n$ on note de façon générale

$$[a, b] = \{x = a + t(b - a) \in \mathbb{R}^n;\ t \in \mathbb{R},\ 0 \leq t \leq 1\}$$

pour désigner l'intervalle fermé de $\mathbb{R}^n$ d'extrémités a et b, et

$$]a, b[= \{x = a + t(b - a) \in \mathbb{R}^n;\ t \in \mathbb{R},\ 0 < t < 1\}$$

pour désigner l'intervalle ouvert de $\mathbb{R}^n$ d'extrémités a et b.

4.1. Formules de Taylor-Lagrange

On rappelle d'abord les formules de Taylor-Lagrange d'ordre un (dite formule des accroissements finis) et d'ordre deux classiques pour les fonctions d'une variable réelle à valeurs réelles.

Proposition 80. *Soit f une fonction définie dans un intervalle fermé $[a, a + h]$ de $\mathbb{R}$ à valeurs dans $\mathbb{R}$.*

1 - (formule des accroissements finis) Si f est continue dans $[a, a + h]$ et dérivable dans $]a, a + h[$, alors il existe $\theta \in]0, 1[$ tel que

$$f(a + h) = f(a) + h\, f'(a + \theta h).$$

2 - Si f est dérivable avec dérivée continue dans $[a, a + h]$ et deux fois dérivable dans $]a, a + h[$, alors il existe $\theta \in]0, 1[$ tel que

$$f(a + h) = f(a) + h\, f'(a) + \frac{1}{2} h^2 f''(a + \theta h).$$

Ces formules se généralisent pour des fonctions de n variables réelles à valeurs réelles de la façon suivante :

Théorème 81. (Formules de Taylor-Lagrange) *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et $[a, a + h]$ un intervalle fermé contenu dans U.*

1 - (ordre un - formule des accroissements finis) Si f est continue dans $[a, a + h]$ et différentiable dans $]a, a + h[$, alors il existe $\theta \in]0, 1[$ tel que

$$f(a + h) = f(a) + df(a + \theta h)\,(h).$$

2 - (ordre deux) Si f est de classe C^1 dans $[a, a + h]$ et deux fois différentiable dans $]a, a + h[$, alors il existe $\theta \in]0, 1[$ tel que

$$f(a + h) = f(a) + df(a)\,(h) + \frac{1}{2} d^2 f(a + \theta h)\,(h, h).$$

Démonstration. On définit la fonction φ dans l'intervalle $[0, 1]$ de $\mathbb{R}$ à valeurs dans $\mathbb{R}$ par

$$\varphi(t) = f(a + th).$$

1 - Si f est continue dans $[a, a+h]$ et différentiable dans $]a, a+h[$, alors par composition la fonction φ est continue dans $[0, 1]$ et dérivable dans $]0, 1[$ de dérivée

$$\varphi'(t) = df(a + th)\,(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + th)h_i.$$

Par la formule des accroissements finis rappelée dans la proposition **??**, il existe alors $\theta \in]0, 1[$ tel que

$$\varphi(1) = \varphi(0) + \varphi'(\theta),$$

c'est-à-dire

$$f(a + h) = f(a) + df(a + \theta h)\,(h).$$

2 - Si f est de classe C^1 dans $[a, a + h]$ et deux fois différentiable dans $]a, a + h[$, alors la fonction φ est dérivable avec dérivée continue dans $[0, 1]$ et deux fois dérivable dans $]0, 1[$ de dérivées

$$\varphi'(t) = df(a + th)\,(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + th)h_i$$

$$\varphi''(t) = d^2 f(a + th)\,(h, h) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a + th)h_i h_j$$

On applique alors la formule de Taylor-Lagrange rappelée précédemment dans la proposition **??** à φ dans $[0, 1]$ pour obtenir la formule pour f. $\diamond$

4.2. Inégalités des accroisements finis

Si maintenant f est une fonction à valeurs dans $\mathbb{R}^p$ avec $p \geq 2$, on peut appliquer la formule des accroissements finis à chaque composante f_j de f pour $j = 1, \ldots, p$ et obtenir un θ_j pour chacune d'elle. Cependant il n'y a pas de raison pour que l'on obtienne le même θ_j pour chacune de ces composantes, autrement dit la formule des accroissements finis ne se généralise pas aux fonctions à valeurs vectorielles dans $\mathbb{R}^p$ pour $p \geq 2$. Ainsi par exemple

Exemple. Pour la fonction $f = (\cos, \sin)$ définie dans $\mathbb{R}$ à valeurs dans $\mathbb{R}^2$, il n'existe pas de θ appartenant à $]0, 1[$ tel que $f(2\pi) - f(0) = 2\pi\, f'(\theta\, 2\pi)$. En effet

$$f(2\pi) - f(0) = (0, 0)$$

alors que

$$f'(x) = (-\sin x, \cos x) \neq (0, 0)$$

pour tout x.

Par contre on a une inégalité des accroissements finis pour les fonctions à valeurs vectorielles :

Théorème 82. (Inégalité des accroissements finis) *Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$ et $[a, a+h]$ un intervalle fermé contenu dans U.*

Si f est continue dans $[a, a+h]$ et différentiable dans $]a, a+h[$, alors il existe $\theta \in]0,1[$ tel que

$$\|f(a+h) - f(a)\| \leq \|df(a+\theta h)(h)\|$$

où $\| \cdot \|$ est une norme quelconque sur $\mathbb{R}^p$.

On peut également établir une inégalité des accroissements finis à l'ordre deux pour contrôler $\|f(a+h) - f(a) - df(a)\,(h)\|$.

Démonstration. Pour y fixé dans $\mathbb{R}^p$, soit ψ la fonction définie dans l'intervalle $[0,1]$ de $\mathbb{R}$ à valeurs dans $\mathbb{R}$ par

$$\psi(t) = < y, f(a+th) > = \sum_{j=1}^{p} y_j\, f_j(a+th).$$

Cette fonction ψ est continue dans $[0,1]$ et dérivable dans $]0,1[$ avec

$$\psi'(t) = < y, df(a+th)(h) > = \sum_{j=1}^{p} y_j \big(\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(a+th) h_i \big).$$

Par la formule des accroissements finis de la proposition **??**, il existe alors un $\theta \in]0,1[$, dépendant de y en général, tel que

$$\psi(1) - \psi(0) = \psi'(\theta)$$

c'est-à-dire

$$< y, f(a+h) - f(a) > = < y, df(a+\theta\, h)(h) > .$$

Pour terminer la démonstration on admet alors le résultat suivant (théorème de Hahn-Banach) : pour tout x dans $\mathbb{R}^p$, il existe un $y \in \mathbb{R}^p$ tel que

$$< y, z > \leq \|z\| \quad \text{pour tout } z \in \mathbb{R}^p \qquad \text{et} \qquad < y, x > = \|x\|.$$

(Dans le cas particulier où la norme $\| \cdot \|$ sur $\mathbb{R}^p$ est la norme euclidienne $\| \cdot \|_2$, alors on vérifie à l'aide de l'inégalité de Cauchy-Schwarz que l'on peut prendre $y = \dfrac{x}{\|x\|_2}$ pour $x \neq 0$)

En choisissant en particulier le y associé à $x = f(a+h) - f(a)$, l'égalité générale précédente

$$< y, f(a+h) - f(a) > = < y, df(a+\theta\, h)(h) >$$

s'écrit alors

$$\|f(a+h) - f(a)\| = < y, df(a+\theta\, h)(h) > .$$

Comme de plus

$$< y, df(a+\theta\, h)(h) > \leq \|df(a+\theta\, h)(h)\|$$

par propriété de y, on obtient l'inégalité des accroissements finis annoncée. $\diamond$

4.3. Formules de Taylor-Young

Théorème 83. *(Formules de Taylor-Young) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et a un point de U.*

1 - Si f est une fois différentiable en un point a de U, alors

$$f(a+h) = f(a) + df(a)\,(h) + \|h\|\,\varepsilon(h),$$

2 - Si f est deux fois différentiable en a, alors

$$f(a+h) = f(a) + df(a)\,(h) + \frac{1}{2} d^2 f(a)\,(h,h) + \|h\|^2\,\varepsilon(h)$$

où dans ces deux formules ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue en 0 et nulle en 0.

Si f est une fonction à valeurs dans $\mathbb{R}^p$ avec $p \geq 2$, on peut appliquer ces formules à chaque composante f_j de f pour $j = 1, \ldots, p$.

Démonstration. 1 - Le premier résultat est la définition de la différentielle d'ordre un.

2 - Si f est deux fois différentiable en a, la fonction f est différentiable dans un voisinage ouvert V de a et ∇f est différentiable en a, par exemple avec $V = \{x \in \mathbb{R}^n;\ \|x-a\| < r\}$ pour un $r > 0$.

Soit alors h fixé dans $\mathbb{R}^n$ tel que $\|h\| < r$.

On considère la fonction r définie dans l'intervalle $[0,1]$ de $\mathbb{R}$ à valeurs dans $\mathbb{R}$ par

$$\begin{aligned} r(t) &= f(a+th) - f(a) - t\,df(a)(h) - \frac{1}{2} t^2\, d^2 f(a)(h,h) \\ &= f(a+th) - f(a) - t < \nabla f(a), h > - \frac{1}{2} t^2 < Hf(a)h, h > . \end{aligned}$$

Cette fonction r est continue dans $[0,1]$ et dérivable dans $]0,1[$ avec

$$r'(t) = < \nabla f(a+th), h > - < \nabla f(a), h > - t < Hf(a)h, h > .$$

D'après la formule des accroissements finis de la proposition **??**, il existe alors un $\theta \in]0,1[$ tel que

$$r(1) - r(0) = r'(\theta),$$

et ainsi

$$f(a+h) - f(a) - df(a)(h) - \frac{1}{2} d^2 f(a)(h,h) = < \nabla f(a+\theta h), h > - < \nabla f(a), h > - \theta < Hf(a)h, h > .$$

Or, comme ∇f est différentiable en a, pour $k \in \mathbb{R}^n$ petit on a

$$\nabla f(a+k) = \nabla f(a) + d(\nabla f)(a)(k) + \|k\|\,\varepsilon_1(k)$$

pour une certaine fonction ε_1 définie dans un voisinage de 0 dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^n$, continue en 0 et nulle en 0. Ainsi pour tout $h \in \mathbb{R}^n$

$$\begin{aligned} < \nabla f(a+k), h > &= < \nabla f(a), h > + < d(\nabla f)(a)(k), h > + \|k\| < \varepsilon_1(k), h > \\ &= < \nabla f(a), h > + < Hf(a)k, h > + \|k\| < \varepsilon_1(k), h > . \end{aligned}$$

Ainsi par application avec $k = \theta h$ pour h petit dans $\mathbb{R}^n$

$$f(a+h) - f(a) - df(a)(h) - \frac{1}{2}d^2 f(a)(h,h) = \|\theta h\| < \varepsilon_1(\theta h), h > .$$

Or en particulier pour la norme euclidienne par exemple, on peut écrire

$$\|\theta h\| < \varepsilon_1(\theta h), h >= \|h\|^2 \varepsilon_2(h)$$

pour une fonction ε_2 définie dans un voisinage de 0 dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, continue en 0 et nulle en 0.

Ceci termine la démonstration. $\diamond$

4.4. Formules de Taylor avec reste intégral

Théorème 84. *(reste intégral) Soit f une fonction définie dans un ouvert U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et $[a, a+h]$ un intervalle fermé contenu dans U.*

1 - Si f est de classe C^1 dans $[a, a+h]$, alors

$$f(a+h) = f(a) + \int_0^1 \big(df(a+th)\,(h)\big)\,dt$$

2 - Si f est de classe C^2 dans $[a, a+h]$, alors

$$f(a+h) = f(a) + df(a)\,(h) + \int_0^1 (1-t)\,\big(d^2 f(a+th)\,(h,h)\big)\,dt$$

Si f est une fonction à valeurs dans $\mathbb{R}^p$ avec $p \geq 2$, on peut appliquer ces formules à chaque composante f_j de f pour $j = 1, \ldots, p$.

Démonstration. On utilise les notations de la démonstration du théorème **??** et on se ramène aux formules de Taylor avec reste intégral d'ordre un et d'ordre deux pour la fonction φ dans $[0, 1]$ à valeurs dans $\mathbb{R}$. $\diamond$

Chapitre 4 : Fonctions implicites - Inversion locale

Pour chaque valeur de x, on s'intéresse aux solutions y de l'équation $f(x, y) = 0$.
On cherche à passer d'une condition implicite entre les variables (x, y) du type $f(x, y) = 0$ à une relation explicite du type $y = \varphi(x)$, pour avoir une équivalence de la forme

$$f(x, y) = 0 \Leftrightarrow y = \varphi(x).$$

C'est l'objet du théorème des fonctions implicites. La fonction φ est souvent appelée la fonction implicite.

Plus particulièrement si $f(x, y) = x - g(y)$, on cherche alors une équivalence de la forme

$$g(y) = x \Leftrightarrow y = \varphi(x).$$

C'est l'objet de l'inversion de la fonction g. La fonction φ est alors appelée la fonction inverse (ou réciproque) de g.

En fait on n'obtiendra en général que des résultats locaux.

Exemple (*d'inversion*) Soit g la fonction définie de $\mathbb{R}$ dans $\mathbb{R}$ par $g(y) = y^2$.
Pour tout $x > 0$, il existe un et un seul $y > 0$ tel que $y^2 = x$ et cette solution y est donnée par $y = \sqrt{x}$. On a l'équivalence

$$\left\{ \begin{array}{l} (x, y) \in]0, +\infty[\times]0, +\infty[\\ y^2 = x \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x \in]0, +\infty[\\ y = \sqrt{x}. \end{array} \right.$$

Ainsi pour tout $a > 0$ il existe un voisinage de a, ici $]0, +\infty[$, tel que la restriction de g à $]0, +\infty[$ soit une bijection de $]0, +\infty[$ sur un voisinage de $g(a)$, ici $]0, +\infty[$, et sa fonction réciproque définie sur $]0, +\infty[$ est donnée par $g^{-1}(x) = \sqrt{x}$ pour $x \in]0, +\infty[$. En fait pour être correct il faut préciser la restriction de g à $]0, +\infty[$ et noter $\left(g_{|]0,+\infty[}\right)^{-1}(x) = \sqrt{x}$ pour $x \in]0, +\infty[$ (pour ne pas confondre avec $\left(g_{|]-\infty,0[}\right)^{-1}(x) = -\sqrt{x}$ pour $x \in]0, +\infty[$). Quand le contexte sera précisé on notera simplement g^{-1}.
Par contre il n'existe pas de voisinage de 0 sur lequel g soit injective puisque pour tout $x > 0$, il existe un $y > 0$ tel que $y^2 = x$ et $(-y)^2 = x$.

Sur cet exemple on remarque deux points :
- on n'a pas d'inversion globale de g dans $\mathbb{R}$ car on ne peut pas inverser g au voisinage de 0,
- on ne peut pas inverser g au voisinage du point 0 où $g'(0) = 0$, c'est-à-dire en un point où la dérivée n'est pas inversible.

On montrera de façon générale que si la différentielle $dg(a)$ au point a d'une fonction g est inversible (en tant qu'application linéaire), alors on peut inverser localement (l'application non linéaire) g au voisinage de a. C'est l'objet du théorème d'inversion locale.

C'est précisément un des buts du calcul différentiel que de déduire une information sur le comportement de g au voisinage de a à partir d'une propriété de $dg(a)$.

En fait on verra ce théorème d'inversion locale comme un corollaire d'un théorème général des fonctions implicites. On donne un exemple avant de donner l'énoncé général.

Exemple (*de fonction implicite.*) Soit f la fonction définie dans $\mathbb{R}^2$ à valeurs dans $\mathbb{R}$ par $f(x,y) = x^2 + y^2 - 1$.

Pour tout $x \in]-1,+1[$ il existe un et un seul $y > 0$ tel que $x^2 + y^2 - 1 = 0$ et cette solution est donnée par $y = \sqrt{1 - x^2}$. Ainsi

$$\begin{cases} (x,y) \in]-1,+1[\times]0,+\infty[\\ x^2 + y^2 - 1 = 0 \end{cases} \Leftrightarrow \begin{cases} x \in]-1,+1[\\ y = \sqrt{1-x^2}. \end{cases}$$

Du point de vue géométrique soit

$$C = \{(x,y) \in \mathbb{R}^2;\ x^2 + y^2 = 1\}$$

le cercle d'équation implicite $x^2+y^2-1 = 0$ et soit (a,b) un point particulier de C avec $a > 0$ et $b > 0$. Le résultat précédent dit donc que ce cercle unité C est représenté localement près de ce point (a,b), plus précisément dans le voisinage $]-1,+1[\times]0,+\infty[$ de ce point, par l'équation

$$y = \sqrt{1-x^2} \qquad \text{pour } \ x \in]-1,+1[$$

c'est-à-dire est le graphe G de la fonction $\varphi(x) = \sqrt{1-x^2}$ dans $]-1,+1[$. Plus précisément

$$C \cap]-1,+1[\times]0,+\infty[= G$$

Par contre on n'a pas de représentation globale de C sous forme d'un graphe.

implicit1-eps-converted-to.pdf

Cet exemple se généralise sous la forme :

Théorème 85 (des fonctions implicites). *Soit U un sous-ensemble ouvert de $\mathbb{R}^n$ pour $n \geq 2$ et f une fonction de classe C^1 dans U à valeurs dans $\mathbb{R}$. On note de façon générale $x = (x', x_n)$ avec $x' = (x_1, \dots, x_{n-1})$.*

Si a est un point de U tel que $f(a) = 0$ et $\dfrac{\partial f}{\partial x_n}(a) \neq 0$, alors il existe un voisinage ouvert V' de a' dans $\mathbb{R}^{n-1}$, un voisinage ouvert V_n de a_n dans $\mathbb{R}$ et une unique fonction φ de *classe C^1 dans V' à valeurs dans V_n tels que*

$$\left\{\begin{array}{l}(x', x_n) \in V' \times V_n \\ f(x', x_n) = 0\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}x' \in V' \\ x_n = \varphi(x').\end{array}\right.$$

De plus pour $i = 1, \ldots, n-1$ et $x' \in V'$

$$\frac{\partial \varphi}{\partial x_i}(x') = -\frac{\dfrac{\partial f}{\partial x_i}(x', \varphi(x'))}{\dfrac{\partial f}{\partial x_n}(x', \varphi(x'))}.$$

On peut également écrire

$$\{x \in V' \times V_n;\ f(x) = 0\} = \{(x', \varphi(x'));\ x' \in V'\}.$$

Ceci implique en particulier que $\varphi(a') = a_n$ et $f(x', \varphi(x')) = 0$ pour $x' \in V'$.

Du point de vue géométrique soit $S = \{x \in U;\ f(x) = 0\}$ l'hypersurface de $\mathbb{R}^n$ (courbe pour $n = 2$, surface pour $n = 3$) d'équation implicite $f(x) = 0$. Le résultat précédent dit donc que S est représentée localement près du point a, plus précisément dans le voisinage $V' \times V_n$ de ce point, par l'équation

$$x_n = \varphi(x') \qquad \text{pour } x' \in V'$$

c'est-à-dire est le graphe G de la fonction φ dans V'. Plus précisément

$$S \cap V' \times V_n = G.$$

implicit2-eps-converted-to.pdf

Démonstration. Pour simplifier on suppose que $n = 2$.

Soit donc U un sous-ensemble de $\mathbb{R}^2$, f une fonction de classe C^1 dans U à valeurs dans $\mathbb{R}$ et (a, b) un point de U tels que par exemple $\dfrac{\partial f}{\partial y}(a, b) > 0$ (quitte à remplacer f par $-f$).

Comme U est un voisinage ouvert de a et que $\dfrac{\partial f}{\partial y}$ est continue et > 0 en (a, b), il existe des voisinages U_a et U_b de a et de b dans $\mathbb{R}$ respectivement tels que

$$U_a \times U_b \subset U \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) > 0 \quad \text{pour } (x, y) \in U_a \times U_b.$$

1 - Existence de φ.

Pour x fixé dans U_a, on note g_x la fonction définie dans U_b par $g_x(y) = f(x, y)$.

Cette fonction g_x est continûment dérivable dans U_b avec

$$g'_x(y) = \frac{\partial f}{\partial y}(x,y) \quad \text{pour} \;\; y \in U_b,$$

et donc g_x est en particulier continue et strictement croissante dans U_b.

Soit alors $\beta > 0$ tel que $[b-\beta, b+\beta] \subset U_b$. Comme $g_a(b) = 0$, alors $g_a(b-\beta) < 0$ et $g_a(b+\beta) > 0$, c'est-à-dire

$$f(a, b-\beta) < 0 \qquad \text{et} \qquad f(a, b+\beta) > 0.$$

Comme f est continue (en x) dans U, il existe donc $\alpha > 0$ tel que $[a-\alpha, a+\alpha] \subset U_a$ et $f(x, b-\beta) < 0$ et $f(x, b+\beta) > 0$, c'est-à-dire

$$g_x(b-\beta) < 0 \quad \text{et} \quad g_x(b+\beta) > 0 \quad \text{pour} \quad x \in [a-\alpha, a+\alpha].$$

Comme g_x est continue et strictement croissante dans $[b-\beta, b+\beta]$, il existe donc un unique point de $[b-\beta, b+\beta]$ noté $\varphi(x)$ tel que $g_x(\varphi(x)) = 0$, c'est-à-dire $f(x, \varphi(x)) = 0$.

implicit3-eps-converted-to.pdf

Ainsi pour tout $x \in]a-\alpha, a+\alpha[$, il existe un unique $\varphi(x) \in]b-\beta, b+\beta[$ tel que $f(x, \varphi(x)) = 0$, et on a l'équivalence

$$\left\{ \begin{array}{l} (x,y) \in]a-\alpha, a+\alpha[\times]b-\beta, b+\beta[\\ f(x,y) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x \in]a-\alpha, a+\alpha[\\ y = \varphi(x), \end{array} \right.$$

On définit ainsi une fonction φ de $]a-\alpha, a+\alpha[$ dans $]b-\beta, b+\beta[$ telle que l'on ait l'équivalence précédente, et cette fonction est unique.

2 - Continuité de φ.

Etant donné $a' \in]a-\alpha, a+\alpha[$ on montre que la fonction φ ainsi définie est continue en a'.

On rappelle que $b-\beta < \varphi(a') < b+\beta$.

Soit alors $0 < \varepsilon \leq \inf(b+\beta-\varphi(a'), -b+\beta+\varphi(a'))$. En particulier comme $g_{a'}$ est strictement croissante sur $[b-\beta, b+\beta]$ et que $g_{a'}(\varphi(a')) = 0$, on a

$$g_{a'}(\varphi(a') - \varepsilon) < 0 \quad \text{car} \quad b-\beta \leq \varphi(a') - \varepsilon < \varphi(a').$$

$$g_{a'}(\varphi(a') + \varepsilon) > 0 \quad \text{car} \quad \varphi(a') < \varphi(a') + \varepsilon \leq b+\beta,$$

c'est-à-dire

$$f(a', \varphi(a') - \varepsilon) < 0 \quad \text{et} \quad f(a', \varphi(a') + \varepsilon) > 0.$$

Comme f est continue dans U, il existe $\alpha' > 0$ tel que $]a'-\alpha', a'+\alpha'[\subset]a-\alpha, a+\alpha[$ et pour $x \in]a'-\alpha', a'+\alpha'[$ on ait $f(x, \varphi(a') - \varepsilon) < 0 \quad \text{et} \quad f(a', \varphi(a') + \varepsilon) > 0$, c'est-à-dire

$$g_x(\varphi(a') - \varepsilon) < 0 \quad \text{et} \quad g_x(\varphi(a') + \varepsilon) > 0.$$

Donc la fonction g_x admet un zéro dans l'intervalle $]\varphi(a') - \varepsilon, \varphi(a') + \varepsilon[$. Or $\varphi(x)$ est le seul zéro de g_x dans l'intervalle $[b-\beta, b+\beta]$ et $]\varphi(a') - \varepsilon, \varphi(a') + \varepsilon[\subset [b-\beta, b+\beta]$. Par conséquent $\varphi(x) \in]\varphi(a') - \varepsilon, \varphi(a') + \varepsilon[$.

implicit4-eps-converted-to.pdf

En conclusion, pour tout $\varepsilon > 0$, il existe $\alpha' > 0$ tel que pour $x \in]a' - \alpha', a + \alpha'[$ on ait $|\varphi(x) - \varphi(a')| < \varepsilon$. Ainsi φ est continue en a'.

3 - Différentiabilité de φ.

Comme f est différentiable en tout point (x, y) de U, on peut écrire pour $(h, k) \in \mathbb{R}^2$ voisin de $(0, 0)$

$$f(x+h, y+k) = f(x, y) + h\frac{\partial f}{\partial x}(x, y) + k\frac{\partial f}{\partial y}(x, y) + (|h| + |k|)\,\varepsilon(h, k)$$

où ε est une fonction définie au voisinage de $(0, 0)$, continue en $(0, 0)$ et nulle en $(0, 0)$.

Soit alors x fixé dans $]a - \alpha, a + \alpha[$. En particulier pour $y = \varphi(x)$ et $k = \varphi(x+h) - \varphi(x)$ pour h suffisamment petit, la relation précédente donne

$$0 = h\frac{\partial f}{\partial x}(x, \varphi(x)) + (\varphi(x+h) - \varphi(x))\frac{\partial f}{\partial y}(x, \varphi(x)) + (|h| + |\varphi(x+h) - \varphi(x)|)\varepsilon(h, \varphi(x+h) - \varphi(x))$$

soit

$$0 = h\Big(\frac{\partial f}{\partial x}(x, \varphi(x)) + \varepsilon_1(h)\Big) + (\varphi(x+h) - \varphi(x))\Big(\frac{\partial f}{\partial y}(x, \varphi(x)) + \varepsilon_2(h)\Big)$$

où

$$\varepsilon_1(h) = \begin{cases} \dfrac{|h|}{h}\,\varepsilon(h, \varphi(x+h) - \varphi(x)) & \text{si } h \neq 0 \\ 0 & \text{si } h = 0 \end{cases}$$

$$\varepsilon_2(h) = \begin{cases} \dfrac{|\varphi(x+h) - \varphi(x)|}{\varphi(x+h) - \varphi(x)}\,\varepsilon(h, \varphi(x+h) - \varphi(x)) & \text{si } \varphi(x+h) - \varphi(x) \neq 0 \\ 0 & \text{si } \varphi(x+h) - \varphi(x) = 0. \end{cases}$$

En particulier

$$|\varepsilon_1(h)| \leq |\varepsilon(h, \varphi(x+h) - \varphi(x))|$$

$$|\varepsilon_2(h)| \leq |\varepsilon(h, \varphi(x+h) - \varphi(x))|$$

donc ε_1 et ε_2 sont des fonctions définies au voisinage de 0, continues en 0 et nulles en 0.

Ainsi comme $\dfrac{\partial f}{\partial y}(x,\varphi(x)) \neq 0$, on peut écrire pour $h \neq 0$ petit

$$\frac{\varphi(x+h)-\varphi(x)}{h} = -\frac{\dfrac{\partial f}{\partial x}(x,\varphi(x)) + \varepsilon_1(h)}{\dfrac{\partial f}{\partial y}(x,\varphi(x)) + \varepsilon_2(h)}$$

d'où l'on en déduit, en faisant tendre h vers 0, $h \neq 0$, que φ est dérivable au point x avec

$$\varphi'(x) = -\frac{\dfrac{\partial f}{\partial x}(x,\varphi(x))}{\dfrac{\partial f}{\partial y}(x,\varphi(x))}\cdot$$

Ceci ayant lieu pour tout $x \in]a-\alpha, a+\alpha[$ et comme φ est continue dans $]a-\alpha, a+\alpha[$, on en déduit que φ est de classe C^1 dans $]a-\alpha, a+\alpha[$. ◇

Remarque. (calcul des dérivées partielles) Sachant que φ est différentiable, on retrouve ses dérivées partielles en partant de l'identité locale

$$f(x',\varphi(x')) = 0 \quad \text{pour} \quad x' \in V'$$

que l'on dérive par rapport à x_i pour $i = 1,\ldots,n-1$. On obtient ainsi

$$\frac{\partial f}{\partial x_i}(x',\varphi(x')) + \frac{\partial f}{\partial x_n}(x',\varphi(x'))\,\frac{\partial \varphi}{\partial x_i}(x') = 0.$$

Comme $\dfrac{\partial f}{\partial x_n}(x',\varphi(x')) \neq 0$ pour $x' \in V'$, on en déduit la dérivée partielle de la fonction φ par rapport à x_i pour $i = 1,\ldots,n-1$

$$\frac{\partial \varphi}{\partial x_i}(x') = -\frac{\dfrac{\partial f}{\partial x_i}(x',\varphi(x'))}{\dfrac{\partial f}{\partial x_n}(x',\varphi(x'))}\cdot$$

Remarque. Le théorème précédent peut être écrit pour les autres variables x_i pour $i = 1,\ldots,n$. Ainsi si $f(a) = 0$ et $\dfrac{\partial f}{\partial x_i}(a) \neq 0$, il existe une fonction φ_i des variables x_j pour $j = 1,\ldots,n$ et $j \neq i$ telle que localement on ait

$$f(x_1,\ldots,x_n) = 0 \Leftrightarrow x_i = \varphi_i(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$$

(avec la convention d'écriture pour $i = 1$ et $i = n$).

Exemple : courbes. Soit U un ouvert de $\mathbb{R}^2$ et f une fonction de classe C^1 de U dans $\mathbb{R}$.
Soit

$$C = \{(x,y) \in U;\ f(x,y) = 0\}$$

la courbe de $\mathbb{R}^2$ d'équation implicite $f(x,y) = 0$.

Si $M_0 = (x_0, y_0)$ est un point de C en lequel $\nabla f(x_0, y_0) \neq 0$, alors la courbe C admet une tangente au point M_0 ayant pour équation

$$(x - x_0)\,\frac{\partial f}{\partial x}(x_0, y_0) + (y - y_0)\,\frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

En effet supposons par exemple que $\dfrac{\partial f}{\partial y}(x_0, y_0) \neq 0$. D'après le théorème des fonctions implicites, cette courbe C d'équation implicite $f(x, y) = 0$ est alors, localement près du point (x_0, y_0), le graphe G d'une fonction φ de classe C^1, c'est-à-dire représentée par l'équation

$$y = \varphi(x).$$

Or ce graphe G a une tangente au point (x_0, y_0) qui a pour vecteur directeur le vecteur $(1, \varphi'(x_0))$. Par conséquent un point (x, y) appartient à cette tangente si et seulement si les vecteurs $(x - x_0, y - y_0)$ et $(1, \varphi'(x_0))$ sont colinéaires, c'est-à-dire si et seulement si

$$\begin{vmatrix} x - x_0 & 1 \\ y - y_0 & \varphi'(x_0) \end{vmatrix} = 0$$

L'équation de cette tangente est donc

$$y - y_0 = \varphi'(x_0)(x - x_0).$$

Comme de plus

$$\varphi'(x) = -\frac{\dfrac{\partial f}{\partial x}(x, \varphi(x))}{\dfrac{\partial f}{\partial y}(x, \varphi(x))}$$

on en déduit le résultat.

On obtiendrait également cette équation de la tangente si on avait $\dfrac{\partial f}{\partial x}(x_0, y_0) \neq 0$ au départ et que l'on avait exprimé implicitement x en fonction de y.

Exemple : surfaces. Soit U un ouvert de $\mathbb{R}^3$ et f une fonction de classe C^1 de U dans $\mathbb{R}$.

Soit

$$S = \{(x, y, z) \in U;\ f(x, y, x) = 0\}$$

la surface de $\mathbb{R}^3$ d'équation implicite $f(x, y, z) = 0$.

Si $M_0 = (x_0, y_0, z_0)$ est un point de S en lequel $\nabla f(x_0, y_0, z_0) \neq 0$, alors la surface admet un plan tangent au point M_0 ayant pour équation

$$(x - x_0)\,\frac{\partial f}{\partial x}(x_0, y_0, z_0) + (y - y_0)\,\frac{\partial f}{\partial y}(x_0, y_0, z_0) + (z - z_0)\,\frac{\partial f}{\partial z}(x_0, y_0, z_0) = 0.$$

En effet supposons par exemple que $\dfrac{\partial f}{\partial z}(x_0, y_0, z_0) \neq 0$. D'après le théorème des fonctions implicites, cette surface S d'équation implicite $f(x, y, z) = 0$ est alors, localement près du point (x_0, y_0, z_0), le graphe G d'une fonction φ de classe C^1, c'est-à-dire représentée par l'équation

$$z = \varphi(x, y).$$

On dit aussi que G est une nappe paramétrée par φ ou d'équation $z = \varphi(x, y)$.

Les vecteurs $(1, 0, \dfrac{\partial \varphi}{\partial x}(x_0, y_0))$ et $(0, 1, \dfrac{\partial \varphi}{\partial y}(x_0, y_0))$ sont indépendants, donc cette nappe a un plan tangent au point M_0 ayant pour vecteurs directeurs ces vecteurs, donc pour équation

$$\begin{vmatrix} x - x_0 & 1 & 0 \\ y - y_0 & 0 & 1 \\ z - z_0 & \dfrac{\partial \varphi}{\partial x}(x_0, y_0) & \dfrac{\partial \varphi}{\partial y}(x_0, y_0) \end{vmatrix} = 0,$$

qui se développe en

$$z - z_0 = (x - x_0)\,\frac{\partial \varphi}{\partial x}(x_0, y_0) + (y - y_0)\,\frac{\partial \varphi}{\partial y}(x_0, y_0).$$

Or l'identité locale $f(x, y, \varphi(x, y)) = 0$ donne par dérivation les expressions des dérivées partielles de φ en (x_0, y_0) en fonction des dérivées partielles de f en (x_0, y_0, z_0). On en déduit le résultat annoncé.

Le théorème des fonctions implicites donné précédemment pour des fonctions à valeurs réelles dans $\mathbb{R}$ se généralise à des fonctions à valeurs vectorielles dans $\mathbb{R}^p$ sous la forme suivante :

Théorème 86 (des fonctions implicites). *Soit U un sous-ensemble ouvert de $\mathbb{R}^n$ et f une fonction de classe C^1 dans U à valeurs dans $\mathbb{R}^p$ avec $1 \le p \le n - 1$.*

Si a est un point de U tel que $f(a) = 0$ et

$$\text{dét}\left[\frac{\partial f_i}{\partial x_j}(a)\right]_{\substack{i=1,\dots,p \\ j=n-p+1,\dots,n}} \neq 0,$$

il existe un voisinage ouvert V de $(a_1, \dots, a_{n-p})$ dans $\mathbb{R}^{n-p}$, un voisinage ouvert W de $(a_{n-p+1}, \dots, a_n)$ dans $\mathbb{R}^p$ et une unique fonction φ de classe C^1 dans V à valeurs dans W tels que

$$\left\{ \begin{array}{l} x \in V \times W \\ f(x) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} (x_1, \dots, x_{n-p}) \in V \\ (x_{n-p+1}, \dots, x_n) = \varphi(x_1, \dots, x_{n-p}). \end{array} \right.$$

Ceci implique en particulier

$$(a_{n-p+1}, \dots, a_n) = \varphi(a_1, \dots, a_{n-p}) \ \text{ et } f(x_1, \dots, x_{n-p}, \varphi(x_1, \dots, x_{n-p})) = 0 \text{ pour } (x_1, \dots, x_{n-p}) \in V.$$

Remarque. Le théorème peut être écrit pour d'autres variables. Ainsi si

$$\text{dét}\left[\frac{\partial f_i}{\partial x_{j_k}}(a)\right]_{\substack{i=1,\dots,p \\ k=1,\dots,p}} \neq 0$$

pour des indices $1 \le j_1 < \dots < j_p \le n$, il existe p fonctions φ_{j_k} des $n - p$ variables x_i pour $i \neq j_l$, $l = 1, \dots, p$ telles que localement

$$f(x) = 0 \Leftrightarrow x_{j_k} = \varphi_{j_k}(x_i, i \neq j_l,\, l = 1, \dots, p) \quad \text{pour} \quad k = 1, \dots, p.$$

Comme corollaire on donne maintenant le théorème d'inversion locale. On va montrer que si une fonction f est différentiable en un point a avec une différentielle $df(a)$ inversible, alors f est inversible localement au voisinage de a.

On va s'intéresser à des fonctions f qui pourront être inversées localement avec les propriétés suivantes :

Définition 87. *Soient V et W des sous-ensembles ouverts de $\mathbb{R}^n$ et f une fonction définie dans V à valeurs dans W. On dit que f est un difféomorphisme de classe C^1 de V sur W si*
- f est de classe C^1 dans V,
- f est une bijection de V sur W, d'inverse notée f^{-1},
- f^{-1} est de classe C^1 dans W.

Théorème 88 (d'inversion locale). *Soit U un sous-ensemble ouvert de $\mathbb{R}^n$ et f une fonction définie dans U à valeurs dans $\mathbb{R}^n$, de classe C^1 dans U.*

Si a est un point de U tel que $df(a)$ soit inversible, alors il existe un voisinage ouvert V de a et un voisinage ouvert W de $f(a)$ dans $\mathbb{R}^n$ tels que la restriction de f à V soit un difféomorphisme de classe C^1 de V sur W.

On peut écrire

$$\{(x, y) \in V \times W;\, y = f(x)\} = \{(x, y) \in \mathbb{R}^n \times W;\, x = f^{-1}(y)\}.$$

Ainsi l'inversibilité de df en a permet donc d'inverser localement la fonction f près de a.

On rappelle que l'application linéaire $df(a)$ est inversible, c'est-à-dire est un isomorphisme de $\mathbb{R}^n$ sur $\mathbb{R}^n$, si et seulement si

$$\text{dét}\left[\frac{\partial f_i}{\partial x_j}(a)\right]_{\substack{i=1,\dots,n\\ j=1,\dots,n}} \neq 0.$$

Démonstration. On se ramène au théorème **??** des fonctions implicites.

On introduit la fonction g définie dans $\mathbb{R}^n \times U$ à valeurs dans $\mathbb{R}^n$ par

$$g(y, x) = y - f(x).$$

Cette fonction g est une fonction de classe C^1 dans $\mathbb{R}^n \times U$ et

$$\text{dét}\left[\frac{\partial g_i}{\partial x_j}(f(a), a)\right]_{\substack{i=1,\dots,n\\ j=1,\dots,n}} = (-1)^n \text{dét}\left[\frac{\partial f_i}{\partial x_j}(a)\right]_{\substack{i=1,\dots,n\\ j=1,\dots,n}} \neq 0$$

D'après le théorème des fonctions implicites, il existe un voisinage W de $f(a)$, un voisinage V de a et une fonction φ de classe C^1 de W à valeurs dans V tels que

$$\begin{cases} (y, x) \in W \times V \\ g(y, x) = 0 \end{cases} \Leftrightarrow \begin{cases} y \in W \\ x = \varphi(y). \end{cases}$$

Cette fonction φ donne donc l'inverse f^{-1} de la (restriction à V de la) fonction f. $\diamond$

Remarque. La règle de dérivation des fonctions composées permet de calculer les dérivées partielles de la fonction inverse dans W sachant qu'elle est a priori différentiable dans W. On part de l'identité

$$(f \circ f^{-1})(y) = y \qquad \text{pour } \ y \in W$$

c'est-à-dire en notant $(\varphi_1, \ldots, \varphi_n)$ les composantes de f^{-1}

$$\begin{cases} f_i(\varphi_1(y), \ldots, \varphi_n(y)) = y_i & \text{pour } \ y \in W \\ i = 1, \ldots, n & \end{cases}$$

et on dérive ces relations par rapport à y_k. On obtient ainsi

$$\begin{cases} \displaystyle\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} \frac{\partial \varphi_j}{\partial y_k} = 0 & i,\, k = 1, \ldots, n, \ \ i \neq k \\ \displaystyle\sum_{j=1}^{n} \frac{\partial f_k}{\partial x_j} \frac{\partial \varphi_j}{\partial y_k} = 1 & k = 1, \ldots, n \end{cases}$$

qui est un système de n équations à n inconnues $\dfrac{\partial \varphi_j}{\partial y_k}$ pour $j = 1, \ldots, n$. Comme son déterminant est le déterminant jacobien

$$\text{dét}\Big[\frac{\partial f_i}{\partial x_j}(f^{-1}(y))\Big]$$

qui est non nul pour $y \in W$, ce système admet une unique solution qui peut par exemple être écrite par les formules de Cramer.

Remarque. On a utilisé le théorème **??** des fonctions implicites pour établir le théorème **??** d'inversion locale. On peut également procéder dans l'autre sens. Ces deux énoncés sont en fait équivalents.

Chapitre 5 - Convexité

La convexité joue un rôle fondamental en optimisation.

1. Ensembles convexes

Définition 89. *Un sous-ensemble U de $\mathbb{R}^n$ est dit*
- convexe si $tx + (1-t)y \in U$ pour x, $y \in U$ et $t \in [0,1]$,
- une variété affine si $tx + (1-t)y \in U$ pour x, $y \in U$ et $t \in \mathbb{R}$.

Autrement dit U est
- convexe si pour tous x et y dans U, l'intervalle fermé $[x,y]$ d'extrémités x et y dans $\mathbb{R}^n$ défini par

$$[x,y] = \{x + t(y-x);\ t \in [0,1]\}$$

est contenu dans U,
- une variété affine si pour tous x et y dans U, la droite affine $d(x; y-x)$ passant par x et de vecteur directeur $y-x$ dans $\mathbb{R}^n$ définie par

$$d(x; y-x) = \{x + t(y-x); t \in \mathbb{R}\}$$

est contenu dans U.

Exemple. Un intervalle de $\mathbb{R}^n$ est convexe. Une boule de $\mathbb{R}^n$ est convexe.

Plus précisément dans le cas de la dimension $n = 1$, on a une caractérisation :

Exemple. Un sous-ensemble U de $\mathbb{R}$ est convexe si et seulement si U est un intervalle.

Exemple. Une variété affine de $\mathbb{R}^n$ est convexe.
Un sous-espace vectoriel de $\mathbb{R}^n$ est une variété affine contenant 0.

Exemple. Soit a et $u \in \mathbb{R}^n$ avec $u \neq 0$ donnés.
La droite affine $d(a; u) = \{a + tu;\ t \in \mathbb{R}\}$ passant par a et de vecteur directeur u est une variété affine, donc convexe.
La demi-droite affine $d_+(a; u) = \{a + tu;\ t \geq 0\}$ passant par a et de vecteur directeur u est convexe.

Exemple. Si une variété affine de $\mathbb{R}^n$ contient une boule, alors elle est identique à $\mathbb{R}^n$.

Exemple. L'intersection d'une famille de sous-ensembles convexes de $\mathbb{R}^n$ est convexe.

Exemple. L'image d'un sous-ensemble convexe de $\mathbb{R}^n$ par une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est convexe dans $\mathbb{R}^p$.

2. Fonctions convexes

Définition 90. *Soit U un sous-ensemble convexe de $\mathbb{R}^n$. Une fonction f de U dans $\mathbb{R}$ est dite*

- convexe si $f(tx+(1-t)y) \leq tf(x)+(1-t)\,f(y)$ pour x, $y \in U$ et $t \in [0,1]$,

- strictement convexe si $f(tx+(1-t)y) < tf(x)+(1-t)\,f(y)$ pour x, $y \in U$, $x \neq y$ et $t \in]0,1[$,

- concave si $-f$ est convexe,

- affine si $f(tx+(1-t)y) = tf(x)+(1-t)\,f(y)$ pour x, $y \in U$ et $t \in [0,1]$.

En dimension $n=1$, une fonction est donc convexe dans un intervalle U de $\mathbb{R}$ si l'intervalle joignant deux points quelconques de son graphe dans $\mathbb{R}^2$ est au-dessus de ce graphe. Elle est strictement convexe si elle est convexe et son graphe ne contient aucun intervalle de longueur > 0. Elle est affine si son graphe est un intervalle, droite ou demi-droite affine.

Remarque. On peut se ramener à des fonctions d'une seule variable : la fonction f est convexe dans U si et seulement si pour tous x et $y \in U$, la fonction g_{xy}, définie dans $[0,1]$ par $g_{xy}(t) = f(tx+(1-t)y)$, est convexe dans $[0,1]$.

Exemple. Une fonction affine dans un convexe U est convexe dans U, mais non strictement convexe dans U.

En fait f est une fonction affine dans un convexe U de $\mathbb{R}^n$ si et seulement s'il existe $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$ tels que $f(x) = < a, x > + b$ pour $x \in U$.

Exemple. Une norme sur $\mathbb{R}^n$ est convexe dans $\mathbb{R}^n$ mais non strictement convexe dans tout $\mathbb{R}^n$.

En effet par l'inégalité triangulaire et par homogénéité, on a pour x, $y \in \mathbb{R}^n$ et $t \in [0,1]$

$$\|tx+(1-t)y\| \leq \|tx\| + \|(1-t)\,y\| = t\|x\| + (1-t)\|\,y\|.$$

Donc la norme $\|\cdot\|$ est convexe dans $\mathbb{R}^n$.

Par contre sa restriction par exemple à une demi-droite $d_+(0;u) = \{tu;\ t \geq 0\}$ de vecteur directeur $u \neq 0$ donné dans $\mathbb{R}^n$, est une fonction affine. En effet pour x et y de la forme $x = ru$ et $y = su$ avec r et $s > 0$, $r \neq s$, on a pour $t \in]0,1[$

$$\begin{aligned} \|tx+(1-t)y\| = \|(tr+(1-t)s)u\| &= (tr+(1-t)s)\|u\| = tr\|u\| + (1-t)s\|u\| \\ &= t\|ru\| + (1-t)\|su\| = t\|x\| + (1-t)\|y\|. \end{aligned}$$

Donc la norme $\|\cdot\|$ n'est pas strictement convexe sur la demi-droite $d_+(0;u)$.

Proposition 91. *Soit U un sous-ensemble convexe et ouvert de $\mathbb{R}^n$ et f une fonction convexe dans U à valeurs dans $\mathbb{R}$. Alors f est continue dans U.*

Démonstration. Soit a un point de U et montrons que f est continue en a.

En remplaçant f par la fonction $x \to f(x+a) - f(a)$ qui est convexe sur le convexe $U - a = \{x - a;\ x \in U\}$, on se ramène au cas où $a = 0$ et $f(0) = 0$.

1 - On montre d'abord que f est majorée dans un voisinage de 0.

Par exemple pour la norme $\| \cdot \|_\infty$ sur $\mathbb{R}^n$, soit $\overline{B}(0,r)$ une boule fermée de centre 0 et de rayon $r > 0$ contenue dans U.

On va montrer qu'il existe M telle que $f(x) \leq M$ pour $x \in B(0, \frac{r}{3})$.

Pour simplifier on se place en dimension $n = 2$.

Etant donné un point x dans $B(0, \frac{r}{3})$, on le décompose en une combinaison convexe de points situés dans $\overline{B}(0,r)$ de la forme

$$x = t_1(r,r) + t_2(-r,0) + t_3(0,-r)$$

où

$$t_1 = \frac{r + x_1 + x_2}{3r} \qquad t_2 = \frac{r - 2x_1 + x_2}{3r} \qquad t_3 = \frac{r - 2x_2 + x_1}{3r}$$

sont des cœfficients ≥ 0 et de somme égale à 1.

Comme f est convexe, on a donc

$$f(x) \leq t_1 f(r,r) + t_2 f(-r,0) + t_3 f(0,-r) \leq |f(r,r)| + |f(-r,0)| + |f(0,-r)|.$$

En posant $M = |f(r,r)| + |f(-r,0)| + |f(0,-r)|$ on a ainsi la majoration annoncée.

2 - On en déduit que f est continue en 0.

Soit ε arbitraire tel que $0 < \varepsilon < s$ où $s = \dfrac{r}{3}$ et soit x tel que $|x| < \varepsilon$.

Si $y = \dfrac{s}{\varepsilon} x$ alors $y \in B(0,s)$ donc $f(y) \leq M$. D'une part de l'égalité

$$x = \frac{\varepsilon}{s} y + (1 - \frac{\varepsilon}{s}) 0$$

on a par la convexité de f

$$f(x) \leq \frac{\varepsilon}{s} f(y) + (1 - \frac{\varepsilon}{s}) f(0)$$

d'où

$$f(x) \leq \frac{\varepsilon}{s} M.$$

D'autre part de l'égalité

$$0 = \frac{s}{s+\varepsilon} x + \frac{\varepsilon}{s+\varepsilon} x'$$

où $x' = -y$, on a par la convexité de f

$$f(0) \leq \frac{s}{s+\varepsilon} f(x) + \frac{\varepsilon}{s+\varepsilon} f(x')$$

d'où

$$0 \leq \frac{s}{s+\varepsilon} f(x) + \frac{\varepsilon}{s+\varepsilon} M$$

soit

$$f(x) \geq -\frac{M}{s} \varepsilon.$$

De ces deux inégalités on déduit que pour tout $\varepsilon \in]0, s[$, on a pour tout $x \in B(0, \varepsilon)$

$$|f(x) - f(0)| \leq \frac{M}{s} \varepsilon.$$

Ceci prouve que f est continue en 0. $\diamond$

Remarque. Une fonction convexe dans un sous-ensemble convexe et ouvert de $\mathbb{R}^n$ y est continue, mais par contre elle n'y est pas nécessairement différentiable. Par exemple la fonction f définie par $f(x) = |x|$ est convexe dans $\mathbb{R}$ mais n'est pas dérivable en 0.

Concernant la stabilité par composition, on a :

Proposition 92. *Soit f une fonction définie dans un sous-ensemble convexe U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et g une fonction définie dans un intervalle I de $\mathbb{R}$ à valeurs dans $\mathbb{R}$ tels que $f(U) \subset I$. Alors*

1 - si f est convexe dans U et g est croissante et convexe dans I, alors $g \circ f$ est convexe dans U,

2 - si f est strictement convexe dans U et g est strictement croissante et strictement convexe dans I, alors $g \circ f$ est strictement convexe dans U.

Démonstration. On montre par exemple la convexité de la fonction $g \circ f$ dans U.

Pour x, $y \in U$ et $t \in [0,1]$ on a par la convexité de f

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

puis par la croissance de g

$$g\big(f(tx + (1-t)y)\big) \leq g(t\, f(x) + (1-t)\, f(y)),$$

et enfin par la convexité de g

$$g(t\, f(x) + (1-t)\, f(y)) \leq t\, g\big(f(x)\big) + (1-t)\, g\big(f(y)\big).$$

Ainsi

$$(g \circ f)(tx + (1-t)y) \leq t\, (g \circ f)(x) + (1-t)\, (g \circ f)(y).$$

Ceci prouve la convexité de la fonction $g \circ f$ dans U. $\diamond$

Exemple. Si p est un réel ≥ 1 et $\|\cdot\|$ une norme sur $\mathbb{R}^n$, la fonction f définie par $f(x) = \|x\|^p$ est convexe dans $\mathbb{R}^n$. Plus précisément la fonction f définie par $f(x) = \|x\|_2^2$, où $\|\cdot\|_2$ est la norme euclidienne, est strictement convexe dans $\mathbb{R}^n$.

Concernant la stabilité par addition et borne supérieure, on a :

Proposition 93. *Soit $(f_i)_{i\in I}$ une famille de fonctions convexes dans un sous-ensemble convexe U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et $(\omega_i)_{i\in I}$ une famille de nombres réels ≥ 0.*

1 - Si la somme $f(x) = \displaystyle\sum_{i\in I} \omega_i f_i(x)$ est définie pour tout $x \in U$, alors la fonction f ainsi définie est convexe dans U.

2 - Si la borne supérieure $f(x) = \displaystyle\sup_{i\in I} f_i(x)$ est définie pour tout $x \in U$, alors la fonction f ainsi définie est convexe dans U.

Démonstration. On montre par exemple la convexité pour la borne supérieure. Soit x, $y \in U$ et $t \in [0,1]$. Pour tout $i \in I$, on a par la convexité de f_i

$$f_i(tx + (1-t)y) \leq t\, f_i(x) + (1-t)\, f_i(y),$$

puis comme $f(x)$ est un majorant de $f_i(x)$ et $f(y)$ est un majorant de $f_i(y)$

$$t\, f_i(x) + (1-t)\, f_i(y) \leq t\, f(x) + (1-t)\, f(y).$$

Donc $t\, f(x) + (1-t)\, f(y)$ est un majorant de chaque $f_i(tx + (1-t)y)$, et par suite

$$f(tx + (1-t)y) \leq t\, f(x) + (1-t)\, f(y). \qquad \diamond$$

Exemple. Soit B un sous-ensemble non vide borné de $\mathbb{R}^n$. Pour $x \in \mathbb{R}^n$ et $b \in B$, on a avec la norme euclidienne

$$| < b, x > | \leq \|b\|_2\, \|x\|_2,$$

donc pour $x \in \mathbb{R}^n$, la famille $\{< b, x >;\, b \in B\}$ est majorée dans $\mathbb{R}$ et non vide, donc admet une borne supérieure. On peut donc poser

$$f(x) = \sup\{< b, x >;\, b \in B\}$$

et la fonction f ainsi définie est convexe dans $\mathbb{R}^n$ car chacune des fonctions f_b définie par $f_b(x) =< b, x >$ est une forme linéaire donc est convexe.

Remarque. La convexité n'est pas conservée en général par la différence, le produit ou le passage à l'enveloppe inférieure (même finie).

Ainsi les fonctions f et g définies dans $\mathbb{R}^2$ par $f(x, y) = x$ et $g(x, y) = y$ sont convexes dans $\mathbb{R}^2$, mais leur produit fg n'est pas convexe.

De même les fonctions f et g définies dans $[-1, 2]$ par $f(x) = x^2$ et $g(x) = (x-1)^2$ sont convexes dans $[-1, 2]$, mais $\inf(f, g)$ n'est pas convexe car son graphe dans $[0, 1]$ est au dessus du segment dont les extrémités ont pour abscisses 0 et 1.

Proposition 94. *Soit f une fonction définie dans un sous-ensemble convexe U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$. On note*

$$Epi(f) = \{(x, t) \in U \times \mathbb{R};\, f(x) \leq t\}$$

son épigraphe et pour $t \in \mathbb{R}$

$$E_t(f) = \{x \in U;\, f(x) \leq t\}$$

son ensemble de niveau t. Alors

1 - la fonction f est convexe dans U si et seulement si son épigraphe est un sous-ensemble convexe de $\mathbb{R}^{n+1}$,

2 - si la fonction f est convexe dans U, ses ensembles de niveau sont des sous-ensembles convexes de $\mathbb{R}^n$ (mais la réciproque n'est pas vraie en général).

Démonstration. On démontre par exemple que si l'épigraphe de f est convexe, alors f convexe dans U.

Pour x et $y \in U$, alors $(x, f(x))$ et $(y, f(y)) \in Epi(f)$ par définition de $Epi(f)$. Donc pour $t \in [0,1]$, le point $t(x, f(x)) + (1-t)(y, f(y)) \in Epi(f)$ par la convexité de $Epi(f)$. Or $t(x, f(x)) + (1-t)(y, f(y)) = (tx + (1-t)y, tf(x) + (1-t)f(y))$, donc par définition de $Epi(f)$ on a

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

ce qui assure la convexité de f. $\diamond$

3. Convexité et différentiabilité d'ordre 1

Proposition 95. *Soit Ω un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction différentiable dans Ω à valeurs dans $\mathbb{R}$ et U un sous-ensemble convexe de Ω. Alors*

1 - les trois propositions suivantes sont équivalentes :

(1-1) la fonction f est convexe dans U,

(1-2) $f(y) - f(x) \geq df(x)\,(y-x)$ pour tous x et $y \in U$,

(1-3) $(df(y) - df(x))\,(y-x) \geq 0$ pour tous x et $y \in U$,

2 - les trois propositions suivantes sont équivalentes :

(2-1) la fonction f est strictement convexe dans U,

(2-2) $f(y) - f(x) > df(x)\,(y-x)$ pour tous x et $y \in U$, $x \neq y$,

(2-3) $(df(y) - df(x))\,(y-x) > 0$ pour tous x et $y \in U$, $x \neq y$.

On rappelle que $df(x)\,(h) =< \nabla f(x), h >$ pour $h \in \mathbb{R}^n$ où $\nabla f(x)$ est le gradient de f en x.

Par exemple en dimension $n = 1$, la propriété (1-2) signifie que quel que soit le point x_0 de l'intervalle U, le graphe de f dans $\mathbb{R}^2$, c'est-à-dire l'arc de courbe d'équation $y = f(x)$, est au-dessus de sa tangente au point $(x_0, f(x_0))$.

En dimension $n = 2$, la propriété (1-2) signifie que quel que soit le point (x_0, y_0) du convexe U, le graphe de f dans $\mathbb{R}^3$, c'est-à-dire la nappe de surface d'équation $z = f(x, y)$ dans $\mathbb{R}^3$ est au-dessus de son plan tangent au point $(x_0, y_0, f(x_0, y_0))$.

Démonstration. 1 - On suppose d'abord que f est convexe dans U.

Soit x et y fixés dans U avec $x \neq y$. Comme Ω est ouvert, on peut définir la fonction g dans un intervalle $]-r, 1+r[$ de $\mathbb{R}$ en posant

$$g(t) = f(x + t(y-x)).$$

Comme f est différentiable en x, la fonction g est dérivable en 0 avec

$$g'(0) = df(x)(y-x).$$

Comme la fonction f est convexe dans U, la fonction g est convexe dans $[0,1]$ et donc

$$g(t) \leq (1-t)\,g(0) + t\,g(1),$$

donc en particulier pour $t > 0$ petit
$$\frac{g(t) - g(0)}{t} \leq g(1) - g(0).$$
Par conséquent à la limite à droite en $t = 0$, on en déduit que
$$g'(0) \leq g(1) - g(0),$$
c'est-à-dire
$$df(x)(y - x) \leq f(y) - f(x).$$

2 - On suppose maintenant plus précisément que f est strictement convexe dans U.

Le raisonnement précédent ne permet pas de conclure puisqu'on ne peut affirmer que l'inégalité stricte
$$\frac{g(t) - g(0)}{t} < g(1) - g(0),$$
est conservée à la limite pour obtenir l'inégalité stricte $g'(0) < g(1) - g(0)$ recherchée.

Soit alors s fixé dans $]0, 1[$. Pour $0 \leq t < s$ on déduit de la convexité de g que
$$g(t) \leq \frac{s - t}{s}\, g(0) + \frac{t}{s}\, g(s),$$
soit
$$\frac{g(t) - g(0)}{t} \leq \frac{g(s) - g(0)}{s}.$$
Or de plus par la convexité stricte de f, on a
$$\frac{g(s) - g(0)}{s} < g(1) - g(0),$$
donc
$$\frac{g(t) - g(0)}{t} \leq \frac{g(s) - g(0)}{s} < g(1) - g(0).$$
Par conséquent, à la limite à droite en $t = 0$, on en déduit que
$$g'(0) \leq \frac{g(s) - g(0)}{s} < g(1) - g(0).$$
Ainsi
$$df(x)(y - x) < f(y) - f(x).$$

3 - Inversement supposons que l'on ait l'inégalité $f(y) - f(x) \geq df(x)(y - x)$ pour tous x et $y \in U$.

Soit x, $y \in U$ avec $x \neq y$ et $t \in]0, 1[$. On a en particulier
$$\begin{array}{l} f(y) \geq f(y + t(x - y)) - t\, df(y + t(x - y))(x - y) \\ f(x) \geq f(y + t(x - y)) + (1 - t)\, df(y + t(x - y))(x - y). \end{array}$$

En multipliant la première inégalité par $1 - t$ et la seconde par t, puis en ajoutant on obtient
$$f(tx + (1 - t)y) \leq t f(x) + (1 - t)\, f(y),$$
ce qui établit la convexité de f.

4 - Avec l'inégalité stricte $f(y) - f(x) > df(x)(y-x)$ pour tous x et $y \in U$, $x \neq y$, on obtient la stricte convexité de f.

Les autres caractérisations sont laissées en exercices. $\diamond$

Proposition 96. *Soit Ω un intervalle ouvert de $\mathbb{R}$, f une fonction dérivable dans Ω à valeurs dans $\mathbb{R}$ et U un intervalle contenu dans U. Alors*

1 - la fonction f est convexe dans U si et seulement si sa dérivée est une fonction croissante dans U,

2 - la fonction f est croissante dans U si et seulement si sa dérivée f' est ≥ 0 dans U.

Démonstration. 1 - C'est une conséquence de l'équivalence entre les propositions (1-2) et (1-3) de la proposition **??**, la proposition (1-3) s'écrivant alors

$$(f'(y) - f'(x))(y-x) \geq 0.$$

2 - Ce résultat est classique. $\diamond$

4. Convexité et différentiabilité d'ordre 2

Proposition 97. *Soit Ω un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction deux fois différentiable dans Ω à valeurs dans $\mathbb{R}$ et U un sous-ensemble convexe de Ω. Alors*

1 - la fonction f est convexe dans U si et seulement si pour tous x et $y \in U$

$$d^2 f(x)\,(y-x, y-x) \geq 0,$$

2 - la fonction f est strictement convexe dans U si pour tous x et $y \in U$ avec $x \neq y$

$$d^2 f(x)\,(y-x, y-x) > 0.$$

On rappelle que $d^2 f(x)\,(h,h) =< Hf(x)\,h, h >$ pour $h \in \mathbb{R}^n$ où $Hf(x)$ est la matrice hessienne de f en x.

Démonstration. 1 - On suppose que f est convexe.

Soit x et $y \in U$ fixés. Par la formule de Taylor-Young d'ordre deux, on a pour tout $t \in [0,1]$

$$f(x+t(y-x)) - f(x) - df(x)(t(y-x)) = \frac{1}{2} d^2 f(x)(t(y-x), t(y-x)) + \|t(y-x)\|^2\, \varepsilon(t(y-x))$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, continue en 0 et nulle en 0. Or, la fonction f étant convexe, la proposition **??** assure que le membre de gauche est positif ou nul. On en déduit que pour tout $t \in]0,1]$

$$0 \leq d^2 f(x)\,(y-x, y-x) + \|y-x\|^2\, \varepsilon(t(y-x)),$$

puis en faisant tendre t vers 0 par valeurs positives, on en déduit que $0 \leq d^2 f(x)\,(y-x, y-x)$ d'après la propriété de la fonction ε en 0.

2 - Inversement soit x et $y \in U$ fixés. Par application de la formule de Taylor-Lagrange à l'ordre deux, il existe $\theta \in]0,1[$ tel que

$$f(y) - f(x) - df(x)\,(y-x) = \frac{1}{2}\,d^2 f(x+\theta(y-x))\,(y-x, y-x)$$

soit en posant $z = x + \theta(y-x)$

$$f(y) - f(x) - df(x)\,(y-x) = \frac{1}{2(1-\theta)^2}\,d^2 f(z)\,(y-z, y-z).$$

Par conséquent si $d^2 f(z)\,(y-z, y-z) \geq 0$ (resp. > 0) pour tous z et $y \in U$ avec $z \neq y$, on en déduit que f est convexe (resp. strictement convexe) dans U d'après la caractérisation de la proposition **??**. $\diamond$

Corollaire 98. *Soit U un sous-ensemble convexe ouvert de $\mathbb{R}^n$ et f une fonction deux fois différentiable dans U à valeurs dans $\mathbb{R}$. Alors*

1 - la fonction f est convexe dans U si et seulement si pour tous $x \in U$ et $h \in \mathbb{R}^n$

$$d^2 f(x)\,(h,h) \geq 0,$$

2 - la fonction f est strictement convexe dans U si pour tous $x \in U$ et $h \in \mathbb{R}^n$ avec $h \neq 0$

$$d^2 f(x)\,(h,h) > 0.$$

Démonstration. Il suffit de noter que si U est ouvert alors pour tous $x \in U$ et $h \in \mathbb{R}^n$ il existe $t \in \mathbb{R}$ tel que $y = x + th \in U$. $\diamond$

Remarque. La condition 2 du corollaire **??** n'est qu'une condition suffisante ; il n'y a pas de réciproque en général. Ainsi par exemple

- la fonction f définie dans $\mathbb{R}$ par $f(x) = x^4$ est strictement convexe dans $\mathbb{R}$, bien que $f''(0) = 0$.

- la fonction f définie dans $\mathbb{R}^2$ par $f(x,y) = x^4 + y^4$ est strictement convexe dans $\mathbb{R}^2$, bien que $Hf(x,y)$ ne vérifie pas la condition de positivité stricte du corollaire **??** quand $x = 0$ ou $y = 0$.

Exemple. La fonction exp est strictement convexe dans $\mathbb{R}$ car $\exp''(x) = \exp(x) > 0$ pour $x \in \mathbb{R}$.

Exemple. La fonction ln est strictement concave dans $]0,+\infty[$ car $\ln''(x) = -\dfrac{1}{x^2} < 0$ pour $x > 0$.

Exemple. (moyenne géométrique) La fonction f définie par $f(x) = (x_1 \cdots x_n)^{\frac{1}{n}}$ dans $U = \{x \in \mathbb{R}^n ; x_i > 0, i = 1, \ldots, n\}$ est concave dans U mais pas strictement concave.

5. Matrices semi-définies positives et définies positives

Les propriétés précédentes de convexité de la fonction f sont donc fondées sur des notions de positivité de la différentielle d'ordre deux $d^2f(x)$, et comme

$$d^2f(x)\,(h,h) =< Hf(x)h, h >$$

où $Hf(x)$ est la matrice hessienne de f en x, ces propriétés de convexité de f sont donc fondées sur des notions de positivité de la matrice $Hf(x)$ dans le sens suivant :

Définition 99. *Une matrice A de taille (n, n) est dite*
1 - semi-définie positive si $< Ax, x > \geq 0$ pour $x \in \mathbb{R}^n$,
2 - définie positive si $< Ax, x >> 0$ pour $x \in \mathbb{R}^n$, $x \neq 0$.
A est dite semi-définie (resp. définie) négative si $-A$ est semi-définie (resp. définie) positive.

En vue de l'application des résultats précédents à des problèmes de convexité, on donne deux caractérisations de la positivité d'une matrice symétrique, tout d'abord à l'aide de ses valeurs propres, ensuite à l'aide de ses déterminants mineurs.

On rappelle que les valeurs propres d'une matrice A de taille (n, n) sont les n zéros $\lambda_1, \ldots, \lambda_n$ réels ou complexes, distincts ou confondus du polynôme caractéristique p_A de la matrice A défini par

$$p_A(\lambda) = \text{dét}\,(A - \lambda\, I).$$

En particulier ces valeurs propres sont réelles quand la matrice est symétrique.

On a alors le résultat suivant qui précise les estimations du corollaire **??** du chapitre 2 :

Proposition 100. *Soit A une matrice symétrique de taille (n, n) et λ_i pour $i = 1, \ldots, n$ ses n valeurs propres (qui sont réelles). Alors*

$$\begin{aligned} \inf_{1\leq i\leq n} \lambda_i &= \inf \left\{ \frac{< Ax, x >}{< x, x >} ; \, x \in \mathbb{R}^n, \, x \neq 0 \right\} \\ \sup_{1\leq i\leq n} \lambda_i &= \sup \left\{ \frac{< Ax, x >}{< x, x >} ; \, x \in \mathbb{R}^n, \, x \neq 0 \right\}. \end{aligned}$$

Les quotients $\dfrac{< Ax, x >}{< x, x >}$ sont appelés les quotients de Rayleigh.

On en déduit la caractérisation suivante :

Corollaire 101. *Soit A une matrice symétrique de taille (n, n). Alors*
1 - A est semi-définie positive si et seulement si ses n valeurs propres sont ≥ 0.
2 - A est définie positive si et seulement si ses n valeurs propres sont > 0.

Comme de plus une matrice A est inversible si et seulement si toutes ses valeurs sont non nulles (car leur produit est égal à $(-1)^n$détA), on en déduit qu'une matrice symétrique semi-définie positive est donc définie positive si et seulement si elle est inversible.

Une autre caractérisation porte sur les déterminants mineurs de A :

Proposition 102. (Règle de Sylvester) *Soit A une matrice symétrique de taille (n, n). Alors A est définie positive si et seulement si les n déterminants des matrices de taille (k, k) obtenues en supprimant les $n-k$ dernières lignes et colonnes de A, pour $k = 1, \ldots, n$, sont > 0.*

Par contre A n'est pas nécessairement semi-définie positive si ces n déterminants mineurs sont ≥ 0.

Enfin on peut utiliser la méthode de Gauss de décomposition de la forme quadratique $< Ax, x >$ en une combinaison linéaire de carrés de formes linéaires

$$< Ax, x >= \sum_{i=1}^{p} \mu_i \Big(\sum_{j=1}^{n} \alpha_{ij} x_j \Big)^2.$$

Exemple. Dans le cas $n = 2$, en vue de l'étude de fonctions de 2 variables, soit une matrice de la forme

$$A = \begin{bmatrix} r & s \\ s & t \end{bmatrix}.$$

Alors, par la règle des valeurs propres, de produit $rt - s^2$ et de somme $r + t$, ou par la méthode de Sylvester, on a :

- A est semi-définie positive si et seulement si $rt - s^2 \geq 0$ et r (ou t) ≥ 0,
- A est semi-définie négative si et seulement si $rt - s^2 \geq 0$ et r (ou t) ≤ 0,
- A est définie positive si et seulement si $rt - s^2 > 0$ et r (ou t) > 0,
- A est définie négative si et seulement si $rt - s^2 > 0$ et r (ou t) < 0.

Chapitre 6 - Optimisation sans contrainte

Soit Ω un sous-ensemble non vide de $\mathbb{R}^n$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$ et U un sous-ensemble de Ω.

Si f est minorée dans U, dans le sens que l'ensemble des valeurs $\{f(x);\, x \in U\}$ est minoré dans $\mathbb{R}$, cet ensemble admet un plus grand minorant appelé sa borne inférieure et noté

$$\inf\{f(x);\, x \in U\} \qquad \text{ou} \qquad \inf_{x\in U} f(x).$$

Le problème de la *minimisation* de f dans U consiste à chercher un point a solution du problème

$$\text{(P)} \qquad a \in U \qquad \text{et} \qquad f(a) = \inf\{f(x);\, x \in U\}.$$

On distingue naturellement les notions locales et globales de solutions dans les sens suivants :

Définition 103. *Soit Ω un sous-ensemble non vide de $\mathbb{R}^n$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$, U un sous-ensemble de Ω et a un point de U. On dit que a est un point de minimum*

1 - global (ou absolu) de f dans U si $f(x) \geq f(a)$ pour tout $x \in U$,

2 - local (ou relatif) de f dans U s'il existe un voisinage V de a dans $\mathbb{R}^n$ tel que $f(x) \geq f(a)$ pour tout $x \in V \cap U$,

3 - global (ou absolu) strict de f dans U si $f(x) > f(a)$ pour tout $x \in U$, $x \neq a$,

4 - local (ou relatif) strict de f dans U s'il existe un voisinage V de a dans $\mathbb{R}^n$ tel que $f(x) > f(a)$ pour tout $x \in V \cap U$, $x \neq a$.

Pour la notion 1 on dit aussi que a est une solution globale du problème (P) ou (par abus de langage) que a est un minimum global de f dans U.

De même pour les notions 2, 3, 4.

Enfin on a écrit ici un problème de minimisation. Tout peut être réécrit de façon similaire pour un problème de maximisation et maximiser une fonction revient à minimiser son opposée.

Par extremum on entendra minimum ou maximum.

Dans ce chapitre 6 on va donner des conditions nécessaires et des conditions suffisantes d'existence d'extremum dans le cas où U est un sous-ensemble *ouvert* général de $\mathbb{R}^n$; les conditions nécessaires portent sur l'équation d'Euler $\nabla f(a) = 0$ et les conditions suffisantes portent sur la positivité de la matrice hessienne $Hf(a)$.

Dans ces problèmes d'optimisation de f dans U, on dit en général que U est l'ensemble des *contraintes*. Mais dans le cas particulier où U est ouvert, on dit souvent, par analogie au cas modèle $U = \mathbb{R}^n$, que le problème d'optimisation est *sans contrainte* et que les extrema éventuels sont *libres*.

Pour terminer ce chapitre, on rappellera le théorème de Weierstrass qui donne une condition suffisante d'existence d'extremum dans le cas où U est un sous-ensemble *fermé* général de $\mathbb{R}^n$.

Dans les chapitres 7 et 8 suivants on s'intéressera à des ensembles de contraintes U particuliers : le cas où U est convexe et le cas où U est de la forme

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \ldots, p\}$$

où les g_i sont des fonctions définies dans Ω à valeurs dans $\mathbb{R}$. Dans ce dernier cas on dit que le problème d'optimisation est *avec contraintes égalité* et que les extrema éventuels sont *liés*.

Enfin dans le chapitre 9 on donnera quelques applications dans le cadre de fonctions f de type quadratique.

Du point de vue terminologie, on dit aussi que le problème d'optimisation (P) est un problème de *programmation*.

1. Optimisation dans un ouvert : conditions nécessaires d'Euler (d'ordre 1) et de Legendre (d'ordre 2)

On donne d'abord la condition nécessaire classique d'existence d'un extremum local portant sur l'annulation des dérivées partielles d'ordre un :

Théorème 104. *(Condition nécessaire : équation d'Euler) Soit U un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U. Si*

1 - f est différentiable en a,

2 - a est un point d'extremum local de f dans U,

alors nécessairement a est un point critique (ou stationnaire) de f, c'est-à-dire vérifie l'équation d'Euler

$$\nabla f(a) = 0.$$

L'équation (vectorielle) d'Euler $\nabla f(a) = 0$ correspond aux n équations (scalaires)

$$\frac{\partial f}{\partial x_i}(a) = 0, \qquad i = 1, \ldots, n.$$

Démonstration. Soit h fixé dans $\mathbb{R}^n$. Par la formule de Taylor-Young d'ordre un, c'est-à-dire la définition de la différentiabilité de f en a, on a pour t réel petit

$$f(a + th) = f(a) + df(a)(th) + \|th\|\, \varepsilon(th)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue en 0 et nulle en 0.

Si a est par exemple un point de minimum local de f dans U, il existe $r > 0$ tel que pour $|t| \leq r$ on ait $f(a + th) \geq f(a)$. Ainsi pour t petit

$$df(a)(th) + \|th\| \, \varepsilon(th) \geq 0.$$

Par suite d'une part pour $t > 0$ petit

$$df(a)(h) + \|h\| \, \varepsilon(th) \geq 0$$

et d'autre part pour $t < 0$ petit

$$-df(a)(h) + \|h\| \, \varepsilon(th) \geq 0.$$

Comme ε est continue en 0 et nulle en 0, on en déduit que $\pm df(a)(h) \geq 0$.

Ainsi $df(a)(h) = 0$. Ceci ayant lieu pour tout $h \in \mathbb{R}^n$, on en déduit que $df(a) = 0$, soit aussi que $\nabla f(a) = 0$. $\diamond$

Remarque. Le fait que U soit ouvert est essentiel. Ainsi par exemple

- pour la fonction f définie dans $\mathbb{R}$ par $f(x) = x$, le point 1 est un point de maximum global dans $[0, 1]$ mais n'est pas un point critique puisque $f'(1) = 1$,

- pour la fonction f définie dans $\mathbb{R}^2$ par $f(x, y) = x + y$, le point $(1, 1)$ est un point de maximum global dans $[0, 1] \times [0, 1]$, mais n'est pas un point critique puisque $\dfrac{\partial f}{\partial x}(1, 1) = 1$.

Remarque. La condition d'Euler est nécessaire, mais non suffisante. Ainsi par exemple

- pour la fonction f définie dans $\mathbb{R}$ par $f(x) = x^3$, le point 0 est un point critique mais n'est pas un point d'extremum puisque $f(-x) < f(0) < f(x)$ pour tout $x > 0$,

- pour la fonction f définie dans $\mathbb{R}^2$ par $f(x, y) = x^2 - y^2$, le point $(0, 0)$ est un point critique de f mais n'est pas un point d'extremum.

Par contre on verra dans le chapitre suivant que si l'ouvert U est de plus convexe et la fonction f convexe, alors la condition nécessaire d'Euler $\nabla f(x) = 0$ est suffisante pour que a soit un point de minimum local et même global de f dans U.

Mise en œuvre pratique du théorème ??. La condition nécessaire d'Euler permet de mettre en évidence les points a de U qui sont d'éventuels points d'extremum de f dans l'ouvert U et pratiquement on commencera par la résolution de l'équation d'Euler

$$\frac{\partial f}{\partial x_i}(a) = 0 \qquad i = 1, \ldots, n.$$

Mais il ne faut pas perdre de vue que cette condition d'Euler n'est qu'une condition nécessaire. Si l'on veut affirmer que les points critiques mis en évidence par cette condition nécessaire correspondent bien à un maximum ou un minimum relatif de la fonction f dans l'ensemble U, et si l'on ne peut pas appliquer les conditions suffisantes que l'on verra plus loin permettant d'affirmer qu'un point critique de f dans U est bien un extremum de f dans U, il faudra compléter cette condition nécessaire par une étude locale directe.

Exemple. On s'intéresse aux extrema dans $\mathbb{R}^2$ de la fonction f définie par $f(x,y) = x^4 + y^4 + 1$.

La fonction f étant différentiable dans $\mathbb{R}^2$, on cherche les extrema éventuels (x,y) dans (l'ouvert) $\mathbb{R}^2$ par la résolution de l'équation d'Euler

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) = 0 \\ \\ \dfrac{\partial f}{\partial y}(x,y) = 0 \end{cases} \qquad \text{soit} \qquad \begin{cases} 4x^3 = 0 \\ \\ 4y^3 = 0. \end{cases}$$

Ainsi f a un seul point critique qui est le point $(0,0)$ et qui est donc le seul point d'extremum éventuel.

Or $f(0,0) = 1$ et $f(x,y) > 1$ pour tout $(x,y) \in \mathbb{R}^2$, $(x,y) \neq (0,0)$. On en déduit que $(0,0)$ est un point de minimum global strict de f dans $\mathbb{R}^2$, et que c'est le seul extremum de f dans $\mathbb{R}^2$. $\diamond$

Exemple. On s'intéresse aux extrema dans $]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[$ de la fonction f définie par $f(x,y) = \sin(x+y)$.

La fonction f étant différentiable dans $\mathbb{R}^2$, on cherche les extrema éventuels (x,y) dans l'ouvert $]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[$ par la résolution de l'équation d'Euler $\cos(x+y) = 0$. Ainsi les points critiques sont les points de $]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[$ tels que $x + y = \frac{\pi}{2} + k\pi$ pour k entier relatif, c'est-à-dire $x + y = \frac{\pi}{2}\cdot$

Or $f(x,y) = 1$ si $x + y = \frac{\pi}{2}$ et $f(x,y) \leq 1$ pour tout (x,y). On en déduit que les points $(x,y) \in]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[$ tels que $x + y = \frac{\pi}{2}$ sont des points de maximum global de f dans $]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[$ et que ce sont les seuls extrema de f dans $]0, \frac{\pi}{2}[\times]0, \frac{\pi}{2}[\cdot$ $\diamond$

Si f est deux fois différentiable, on peut ajouter à cette première condition nécessaire d'ordre 1 d'Euler, une deuxième condition nécessaire d'ordre deux portant sur la positivité de la matrice hessienne.

Proposition 105. *(Condition nécessaire de Legendre) Soit U un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U. Si*

1 - f est deux fois différentiable en a,

2 - a est un point de minimum (resp. maximum) local de f dans U,

alors nécessairement

1 - a est un point critique de f, c'est-à-dire vérifie l'équation d'Euler

$$\nabla f(a) = 0,$$

2 - $Hf(a)$ est une matrice semi-définie positive (resp. négative), c'est-à-dire vérifie la condition de Legendre

$$< Hf(a)h, h > \geq 0 \;\; (resp. \leq 0) \quad pour \quad h \in \mathbb{R}^n.$$

Démonstration. Soit h fixé dans $\mathbb{R}^n$. Par la formule de Taylor-Young d'ordre deux, on a pour t réel

$$f(a+th) = f(a) + df(a)(th) + \frac{1}{2}d^2f(a)(th,th) + \|th\|^2\,\varepsilon(th)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue en 0 et nulle en 0.

Si a est par exemple un point de minimum local de f dans U, il existe $r > 0$ tel que pour $|t| \leq r$ on ait $f(a+th) \geq f(a)$; de plus $df(a)(th) = 0$ par la condition nécessaire d'ordre un précédente, donc

$$\frac{1}{2}d^2f(a)(th,th) + \|th\|^2\,\varepsilon(th) \geq 0.$$

Par suite

$$\frac{1}{2}d^2f(a)(h,h) + \|h\|^2\,\varepsilon(th) \geq 0,$$

et comme ε est continue en 0 et nulle en 0, on en déduit que $d^2f(a)(h,h) \geq 0$, soit aussi $< Hf(a)h, h > \geq 0$. ◇

Remarque. La condition nécessaire que l'on vient de mettre en évidence n'est pas suffisante pour l'existence d'un extremum local. Par exemple

- pour la fonction f définie dans $\mathbb{R}$ par $f(x) = x^3$, on a $f'(0) = 0$ et $f''(0) = 0$; cependant le point 0 n'est pas un extremum local,

- pour la fonction f définie dans $\mathbb{R}^2$ par $f(x,y) = x^3 + y^2$, le point $(0,0)$ est un point critique puisque $\dfrac{\partial f}{\partial x}(0,0) = \dfrac{\partial f}{\partial y}(0,0) = 0$ et de plus $Hf(0,0)$ est semi-définie positive puisque

$$Hf(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

Cependant le point $(0,0)$ n'est pas un extremum local pour f dans $\mathbb{R}^2$ puisque par exemple $f(-x,0) < f(0,0) < f(x,0)$ pour tout $x > 0$.

2. Optimisation dans un ouvert : conditions suffisantes fortes de Legendre

Compte-tenu de la remarque précédente, on est donc conduit, pour obtenir une condition suffisante pour l'existence d'un extremum local, à faire une hypothèse plus forte sur la semi-positivité de la matrice hessienne de f dans un voisinage du point a (cf. proposition **??**) ou sur la positivité de la matrice hessienne de f au point a (cf. proposition **??**).

Proposition 106. *(Condition suffisante forte de Legendre 1) Soit U un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U. S'il existe un voisinage ouvert V de a contenu dans U tel que*

1 - f est deux fois différentiable dans V,

2 - $\nabla f(a) = 0$,

3 - $Hf(x)$ est une matrice semi-définie positive (resp. négative) pour $x \in V$,

alors a est un point de minimum (resp. maximum) local de f dans U.

Démonstration. V contient une boule de centre a.

Par la formule de Taylor-Lagrange d'ordre deux, pour tout h dans cette boule il existe $\theta \in]0,1[$ tel que

$$f(a+h) = f(a) + df(a)(h) + \frac{1}{2}d^2f(a+\theta h)(h,h).$$

Par suite, pour h dans cette boule, on a

$$f(a+h) - f(a) \geq 0$$

et donc a est un point de minimum local de f dans U. ◇

Exemple. Soit la fonction f définie dans $]0,\frac{\pi}{2}[\times]0,\frac{\pi}{2}[$ par $f(x,y) = \sin(x+y)$. Ses points critiques sont tels que $x+y=\frac{\pi}{2}$ et de plus

$$Hf(x,y) = -\sin(x+y)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

est semi-définie négative pour tout $(x,y) \in]0,\frac{\pi}{2}[\times]0,\frac{\pi}{2}[$.

On retrouve ainsi que les points critiques $(x,y) \in]0,\frac{\pi}{2}[\times]0,\frac{\pi}{2}[$ tels que $x+y=\frac{\pi}{2}$ sont des points de maximum local de f dans $]0,\frac{\pi}{2}[\times]0,\frac{\pi}{2}[$ et comme de plus $f(x,y) \leq 1$ que ce sont des points de maximum global. ◇

Proposition 107. *(Condition suffisante forte de Legendre 2) Soit U un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction définie dans U à valeurs dans $\mathbb{R}$ et a un point de U. Si*

1 - f est deux fois différentiable en a,

2 - $\nabla f(a) = 0$,

3 - $Hf(a)$ est une matrice définie positive (resp. négative),

alors a est un point de minimum (resp. maximum) local strict de f dans U.

Démonstration. Comme on l'a vu dans le corollaire **??** du chapitre 2, si la matrice $Hf(a)$, c'est-à-dire aussi la forme quadratique associée $d^2f(a)(h,h) =< Hf(a)h, h >$, est définie positive, alors il existe $m > 0$ telle que pour tout $h \in \mathbb{R}^n$

$$d^2f(a)(h,h) \geq m\,\|h\|^2.$$

Par la formule de Taylor-Young d'ordre deux, on a pour h de norme petite

$$f(a+h) = f(a) + df(a)(h) + \frac{1}{2}d^2f(a)(h,h) + \|h\|^2\,\varepsilon(h)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue en 0 et nulle en 0, et par suite

$$f(a+h) - f(a) \geq (\frac{m}{2} - |\varepsilon(h)|)\|h\|^2.$$

Or par la propriété de ε en 0, il existe $\delta > 0$ tel que $|\varepsilon(h)| \leq \frac{m}{4}$ pour $\|h\| \leq \delta$.

Ainsi pour $\|h\| \leq \delta$ on a

$$f(a+h) - f(a) > 0$$

et donc a est un point de minimum local strict de f dans U. $\diamond$

Exemple. Soit la fonction f définie dans $\mathbb{R}^2$ par $f(x,y) = x^2 + y^2 - 2x$. Son seul point critique est le point $(1,0)$ et de plus

$$Hf(1,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

qui est définie positive.

Donc le point critique $(1,0)$ est un point de minimum local strict de f dans $\mathbb{R}^2$, et comme $f(x,y) = (x-1)^2 + y^2 - 1 \geq -1 = f(1,0)$, il est global. Enfin comme il n'y a qu'un point critique, il est le seul extremum de f dans $\mathbb{R}^2$. $\diamond$

Remarque. Il n'existe pas de réciproque à ces deux résultats. Ainsi par exemple le point $(0,0)$ est un point critique de la fonction f définie dans $\mathbb{R}^2$ par $f(x,y) = x^4 + y^4$ et la matrice hessienne $Hf(0,0)$ est la matrice nulle (donc non définie positive). Cependant ce point $(0,0)$ est un point de minimum global strict de f dans $\mathbb{R}^2$ car $f(0,0) < f(x,y)$ pour tout $(x,y) \neq (0,0)$.

On rappelle que la matrice hessienne $Hf(a)$ (qui est symétrique) est définie positive (resp. négative) si et seulement si ses n valeurs propres sont > 0 (resp. < 0). Dans ce cas si a est de plus un point critique de f, alors il est un point de minimum (resp. maximum) local de f dans U. Par contre si le point a est un point critique et si la matrice hessienne $Hf(a)$ n'a pas de valeur propre nulle, mais a au moins une valeur propre > 0 et une valeur propre < 0, alors a n'est pas un extremum local, mais un point appelé *point-col* ou *point-selle* local.

Plus précisément dans le cas de la dimension $n = 2$, on peut écrire :

Corollaire 108. *(Cas n=2) Soit U un sous-ensemble ouvert de $\mathbb{R}^2$ et f une fonction définie dans U à valeurs dans $\mathbb{R}$, deux fois différentiable en un point* **critique** *a de U. On pose*

$$\frac{\partial^2 f}{\partial x^2}(a) = r, \quad \frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a) = s, \quad \frac{\partial^2 f}{\partial y^2}(a) = t.$$

1 - Si $rt - s^2 > 0$, alors a est un point d'extremum local strict de f dans U, minimum si $r > 0$ (ou $t > 0$) et maximum si $r < 0$ (ou $t < 0$).

2 - Si $rt - s^2 < 0$, alors a n'est pas un point d'extremum local de f dans U mais un point col local de f dans U : il existe h_1 et $h_2 \in \mathbb{R}^n$ avec $< h_1, h_2 >= 0$ tels que pour $t \in \mathbb{R}$, $t \neq 0$ petit on ait $f(a + th_1) < f(a) < f(a + th_2)$.

3 - Si $rt - s^2 = 0$, alors on ne peut rien dire de général pour a.

Démonstration. Le nombre $rt - s^2$ est le produit des deux valeurs propres λ_1 et λ_2 de la matrice hessienne $Hf(a)$.

1 - La conclusion 1 est une conséquence de la proposition **??** et de la caractérisation des matrices définies positives de taille $(2, 2)$.

2 - Soit h_1 et h_2 des vecteurs propres associés aux valeurs propres $\lambda_1 < 0$ et $\lambda_2 > 0$ respectivement avec $< h_i, h_i >= 1$ si $i = 1, 2$ et $< h_1, h_2 >= 0$. Par la formule de Taylor-Young d'ordre deux, on a pour t réel petit

$$f(a + th_i) = f(a) + \frac{t^2}{2}\lambda_i + t^2\,\varepsilon(th_i)$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^2$, continue en 0 et nulle en 0.

Ainsi a n'est pas un point d'extremum local de f dans U puisque $f(a + th_1) < f(a) < f(a + th_2)$ pour $t \in \mathbb{R}$, $t \neq 0$ petit. Ce type de point est appelé un point col local de f dans U.

3 - Si $rt - s^2 = 0$, on ne peut rien dire de général pour a. Ainsi par exemple

- si f est la fonction définie dans $\mathbb{R}^2$ par $f(x, y) = x^4 - y^2$, le point $(0, 0)$ est un point critique en lequel $rt - s^2 = 0$. Ce point n'est pas un extremum local mais un point col puisque $f(0, y) < f(0, 0) < f(x, 0)$ pour tous x et $y \in \mathbb{R}$.

- si f est la fonction définie dans $\mathbb{R}^2$ par $f(x, y) = x^4 + y^4$, le point $(0, 0)$ est point critique en lequel $rt - s^2 = 0$. Ce point est un minimum global strict car $f(0, 0) < f(x, y)$ pour $(x, y) \neq (0, 0)$. ◇

Du point de vue géométrique soit S la nappe de $\mathbb{R}^3$ d'équation $z = f(x, y)$ pour $(x, y) \in U$ et soit a un point critique de f. Le plan d'équation $z = f(a)$ est tangent à S au point $(a, f(a))$, et au voisinage de S

- si $rt - s^2 > 0$ et $r > 0$, alors S est située au-dessus du plan tangent,
- si $rt - s^2 > 0$ et $r < 0$, alors S est située au-dessous du plan tangent,
- si $rt - s^2 < 0$, alors S traverse le plan tangent,
- si $rt - s^2 = 0$, alors on ne peut rien dire de général.

Exemple. On s'intéresse aux extrema dans $U = \{(x, y) \in \mathbb{R}^2 ;\ x > 0,\ y > 0\}$ de la fonction f définie dans U par $f(x, y) = x + y + \dfrac{2}{xy}\cdot$

Comme U est ouvert, on cherche les extrema éventuels (x, y) de f dans U par la résolution de l'équation d'Euler

$$\begin{cases} \dfrac{\partial f}{\partial x}(x, y) = 1 - \dfrac{2}{x^2 y} = 0 \\ \dfrac{\partial f}{\partial x}(x, y) = 1 - \dfrac{2}{x y^2} = 0 \end{cases}$$

Ainsi f a un seul point critique qui est le point $a = (\sqrt[3]{2}, \sqrt[3]{2})$.

Or avec les notations du corollaire $r = 2^{\frac{2}{3}}$, $s = 2^{-\frac{1}{3}}$, et $t = 2^{\frac{2}{3}}$, donc $rt - s^2 > 0$ et $r > 0$.

On en déduit que $(\sqrt[3]{2}, \sqrt[3]{2})$ est un point de minimum local strict de f dans $\mathbb{R}^2$, et que c'est le seul extremum de f dans $\mathbb{R}^2$.

Résumé : mise en œuvre pratique de la recherche d'un extremum de f dans **l'ouvert** U :

1 - on cherche les éventuels extrema a par la résolution de l'équation d'Euler $\nabla f(a) = 0$ (condition nécessaire du théorème **??**),

2 - on cherche si ces points critiques sont des extrema :

2-1 - soit par la positivité de la matrice hessienne (conditions suffisantes des propositions **??** et **??**, et corollaire **??** si n=2),

2-2 - soit par une étude directe locale du signe de $f(x) - f(a)$.

3. Optimisation dans un fermé : condition suffisante de Weierstrass

L'un des plus anciens théorèmes d'existence d'une solution d'un problème d'optimisation est le théorème de Weierstrass déjà démontré dans les corollaires **??** et **??** du chapitre 2 :

Théorème 109 (**Théorème de Weierstrass).** *Soit U un sous-ensemble non vide de $\mathbb{R}^n$ et f une fonction définie dans U à valeurs dans $\mathbb{R}$. Si*

1 - U est fermé dans $\mathbb{R}^n$,

2 - f est continue dans U et de plus coercive lorsque U n'est pas borné,

alors il existe au moins un point de minimum global de f dans U.

Autrement dit $f(U)$ est minoré dans $\mathbb{R}$ et il existe au moins un $a \in U$ tel que

$$f(a) = \inf \{f(x);\, x \in U\}.$$

Démonstration. Dans le cadre de la minimisation qui nous intéresse ici, on donne une démonstration (légérement) différente de celle donnée pour le corollaire **??** du chapitre 2 et reposant sur la notion de suite minimisante, dans le cas où U est fermé et borné dans $\mathbb{R}^n$.

1 - Si le sous-ensemble $f(U)$ n'est pas minoré dans $\mathbb{R}$, il existe une suite $(x_k)_k$ de points de U telle que $f(x_k) \leq -k$ pour tout k. Comme U est fermé et borné, il existe une sous-suite $(x_{k_j})_j$ qui converge vers un point x de U. Comme f est continue en x, la suite $(f(x_{k_j}))_j$ converge vers $f(x)$, donc est bornée, en particulier minorée. Ce qui est en contradiction avec l'hypothèse faite sur la suite $(f(x_{k_j}))_j$.

2 - L'ensemble $f(U)$ étant non vide et minoré, il admet une borne inférieure. Soit $(x_k)_k$ une suite minimisante pour f dans U, c'est-à-dire une suite de points de U telle que $\lim\limits_{k\to\infty} f(x_k) = \inf \{f(x);\, x \in U\}$, cette suite existant par définition de la borne inférieure $\inf \{f(x);\, x \in U\}$.

Comme U est fermé et borné, il existe une sous-suite $(x_{k_j})_j$ qui converge vers un élément a de U. Comme f est continue en a, la suite $(f(x_{k_j}))_j$ converge vers $f(a)$.

Or $\lim\limits_{j\to\infty} f(x_{k_j}) = \lim\limits_{k\to\infty} f(x_k)$, donc $f(a) = \inf \{f(x);\, x \in U\}$. $\diamond$

On a un résultat correspondant pour la maximisation de f en passant à $-f$.

On va s'intéresser dans les deux chapitres suivants à des contraintes particulières. Auparavant pour terminer ce chapitre on fait la remarque suivante :

Remarque (générale concernant la répartition d'extrema éventuels).

Tout d'abord si U est un sous-ensemble quelconque de $\mathbb{R}^n$, la réunion de tous les ouverts de $\mathbb{R}^n$ contenus dans U est le plus grand ouvert de $\mathbb{R}^n$ contenu dans U ; il est noté $\overset{\circ}{U}$ et appelé intérieur de U. Il peut être vide.

Les points de U en lesquels la fonction f peut présenter un extremum sont donc à chercher parmi

1 - les points du bord $U \setminus \overset{\circ}{U}$, (ce sera en général un problème d'optimisation avec contraintes d'égalité tel qu'on le verra dans le chapitre 8),

2 - les points de l'intérieur $\overset{\circ}{U}$ en lesquels f est différentiable et qui sont critiques pour f (c'est un problème d'optimisation dans un ouvert tel qu'on vient de le voir),

3 - les points de l'intérieur $\overset{\circ}{U}$ en lesquels f n'est pas différentiable.

Si on est dans les conditions particulières d'application du théorème de Weierstrass pour lesquelles il existe effectivement au moins un point de minimum global de f dans U et au moins un point de maximum global, on pourra ensuite comparer les différentes valeurs prises aux différents points candidats trouvés, ce qui permettra de déterminer le ou les points qui donnent les extrema globaux de la fonction dans U.

On donne un exemple simple pour illustrer cette remarque générale :

Exemple. On cherche les extrema dans $U = \{(x, y) \in \mathbb{R}^2;\ x^2 + y^2 \leq 2\}$ de la fonction f définie dans $\mathbb{R}^2$ par $f(x, y) = 2x^2 + y^2 + 1$.

L'ensemble U étant fermé et borné dans $\mathbb{R}^2$ (car U est la boule fermée de centre 0 et de rayon $\sqrt{2}$ pour la norme euclidienne) et f étant continue dans $\mathbb{R}^2$ (car polynomiale), le théorème de Weierstrass assure que cette fonction f admet au moins un minimum global et un maximum global dans U.

On cherche les candidats éventuels d'abord sur le bord qui est la sphère de centre 0 et de rayon $\sqrt{2}$, ensuite dans l'intérieur qui est la boule ouverte de centre 0 et de rayon $\sqrt{2}$.

Sur le bord $x^2 + y^2 = 2$, on se ramène dans ce cas particulier à un problème en une seule variable car $f(x, y) = x^2 + 3$ pour $x \in [-\sqrt{2}, +\sqrt{2}]$. Ainsi les points $(-\sqrt{2}, 0)$ et $(+\sqrt{2}, 0)$ sont des points de maximum de f sur le bord de U avec $f(\pm\sqrt{2}, 0) = 5$, et les points $(0, -\sqrt{2})$ et $(0, +\sqrt{2})$ sont des points de minimum de f sur le bord de U avec $f(0, \pm\sqrt{2}) = 3$.

A l'intérieur, la fonction f étant différentiable, les extrema éventuels sont à chercher parmi les points critiques de f. Or $(0, 0)$ est le seul point critique et $f(x, y) > f(0, 0) = 1$ pour $(x, y) \in U$, $(x, y) \neq (0, 0)$.

En conclusion les points $(-\sqrt{2}, 0)$ et $(+\sqrt{2}, 0)$ sont les deux points de maximum global de f dans U et $(0, 0)$ est le seul point de minimum global de f dans U, et il est strict.

Chapitre 7 - Optimisation convexe

La convexité joue un rôle crucial et apporte des propriétés importantes dans l'étude des problèmes d'optimisation.

On s'intéresse ici à des problèmes d'optimisation pour lesquels l'ensemble des contraintes U est convexe. Dans ce cas on donne une condition nécessaire par l'inéquation d'Euler, cette condition nécessaire devenant suffisante si de plus la fonction f est aussi convexe.

Dans ce qui suit les énoncés seront donnés dans le cadre de fonctions convexes et de minima, mais on a des énoncés correspondant dans le cadre de fonctions concaves et de maxima.

1. Optimisation dans un convexe : condition nécessaire d'Euler (d'ordre 1)

Proposition 110. *(Condition nécessaire pour U convexe : inéquations d'Euler) Soit Ω un sous-ensemble ouvert de $\mathbb{R}^n$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$, U un sous-ensemble convexe de Ω et a un point de U. Si*

1 - f est différentiable en a,

2 - a est un point de minimum local de f dans U,

alors a vérifie nécessairement les inéquations d'Euler

$$< \nabla f(a), x - a > \geq 0 \quad \textit{pour} \quad \textit{tout} \ \ x \in U.$$

Démonstration. Soit x un point quelconque de U qu'on écrit $x = a + h$. L'ensemble U étant convexe, les points $a+th$ appartiennent à U pour $t \in [0, 1]$. Comme f est différentiable en a, on peut écrire

$$f(a + th) - f(a) = df(a)\,(th) + \|th\|\,\varepsilon(th) = t(df(a)\,(h) + \|h\|\,\varepsilon(th))$$

où ε est une fonction définie dans un voisinage de 0 dans $\mathbb{R}^n$, continue en 0 et nulle en 0.

Alors $df(a)\,(h)$ est nécessairement ≥ 0, sans quoi la différence $f(a + th) - f(a)$ serait < 0 pour $t > 0$ suffisamment petit.

Ainsi nécessairement $df(a)\,(x - a) \geq 0$ pour $x \in U$. ◇

Remarque. Si le sous-ensemble U est un sous-espace vectoriel, les inéquations d'Euler deviennent les équations d'orthogonalité $< \nabla f(a), x >= 0$ pout tout $x \in U$.

Remarque. Si le sous-ensemble convexe U est ouvert, les inéquations d'Euler deviennent les équations $< \nabla f(a), x >= 0$ pout tout $x \in \mathbb{R}^n$, c'est-à-dire $\nabla f(a) = 0$; on retrouve bien évidemment l'équation d'Euler qui est une condition nécessaire pour U ouvert quelconque.

2. Optimisation pour une fonction convexe dans un convexe : condition nécessaire et suffisante d'Euler (d'ordre 1)

Si *de plus* la fonction f est convexe, la condition nécessaire précédente devient suffisante. On va d'abord montrer que dans ce cas on peut aussi s'affranchir du caractère local de l'extremum et obtenir l'unicité de l'extremum.

Proposition 111. *Soit U un sous-ensemble convexe non vide de $\mathbb{R}^n$ et f une fonction convexe définie dans U à valeurs dans $\mathbb{R}$.*

1 - Si a est un point de minimum local de f dans U, alors a est un point de minimum global de f dans U.

2 - Si f est strictement convexe dans U, alors il existe au plus un point de minimum global de f dans U (qui est donc un minimum global strict s'il existe).

3 - S'il n'est pas vide, l'ensemble des minima globaux de f dans U est un sous-ensemble convexe de U.

Démonstration. 1 - Soit a un point de minimum local de f dans U.

Soit x un point quelconque de U qu'on écrit sous la forme $x = a+h$. D'après la convexité de f, on a pour $0 \leq t \leq 1$

$$f(a+th) \leq (1-t)\, f(a) + t\, f(a+h),$$

soit

$$f(a+th) - f(a) \leq t(f(a+h) - f(a)).$$

Comme a est un point de minimum local de f dans U, il existe $t > 0$ tel que $f(a) \leq f(a+th)$, ce qui implique que

$$0 \leq f(a+h) - f(a).$$

Ceci ayant lieu pour $x = a+h$ quelconque dans U, c'est que a est un point de minimum global de f dans U.

2 - Si a et $a' \in U$ sont deux points distincts de minimum global de f dans U, alors pour $0 < t < 1$ le point $ta + (1-t)a' \in U$ et vérifie

$$f(ta + (1-t)a') < tf(a) + (1-t)f(a') = f(a)$$

ce qui contredit la définition de a comme point de minimum global de f dans U.

3 - Si a et $a' \in U$ sont deux points de minimum global de f dans U, alors pour $0 \leq t \leq 1$ le point $ta + (1-t)a' \in U$ et vérifie

$$f(ta + (1-t)a') \leq f(a) + (1-t)f(a') = f(a)$$

Donc nécessairement $f(ta + (1-t)a') = f(a)$ et par suite le point $a + (1-t)a'$ est aussi un point de minimum global de f dans U. $\diamond$

Dans le cas différentiable, on peut caractériser les points de minimum de la façon suivante :

Corollaire 112. *(Condition nécessaire et suffisante pour U convexe et f convexe : inéquations d'Euler) Soit Ω un sous-ensemble ouvert de $\mathbb{R}^n$, U un sous-ensemble convexe de Ω, f une fonction convexe définie dans Ω à valeurs dans $\mathbb{R}$ et a un point de U.*

Si f est différentiable en a, les trois propositions suivantes sont équivalentes :

1 - a est un point de minimum local de f dans U,

2 - a est un point de minimum global de f dans U,

3 - a vérifie les inéquations d'Euler $< \nabla f(a), x - a > \geq 0$ pour tout $x \in U$.

Démonstration. 1 - Si a est un point de minimum local de f dans U, alors il est un point de minimum global de f dans U d'après la proposition **??**.

2 - Si a est un point de minimum global de f dans U, alors il vérifie nécessairement les inéquations d'Euler d'après la proposition **??** (sans supposer f convexe).

3 - Inversement si f est convexe dans U et différentiable en a, alors nécessairement pour $x \in U$ on a

$$f(x) - f(a) \geq < \nabla f(a), x - a >$$

comme on l'a montré dans la démonstration de la condition nécessaire de la proposition **??** du chapitre 5.

Par suite si $< \nabla f(a), x - a > \geq 0$ pour tout $x \in U$, alors a est un point de minimum global de f dans U.

Plus précisément si U est de plus ouvert, on peut caractériser le point de minimisation par l'équation d'Euler :

Corollaire 113. *(Condition nécessaire et suffisante pour U ouvert convexe et f convexe : équation d'Euler) Soit U un sous-ensemble convexe ouvert de $\mathbb{R}^n$, f une fonction convexe définie dans U à valeurs dans $\mathbb{R}$ et a un point de U. Si f est différentiable en a, les trois propositions suivantes sont équivalentes :*

1 - a est un point de minimum local de f dans U,

2 - a est un point de minimum global de f dans U,

3 - a est un point critique de f, c'est-à-dire vérifie l'équation d'Euler $\nabla f(a) = 0$.

On peut traduire cette propriété localement : si une fonction f définie dans un ouvert U est convexe (resp. concave) dans une boule centrée en un point *critique* a, alors elle admet un minimum (resp. maximum) local en ce point a. En particulier pour $n = 2$, la surface de $\mathbb{R}^3$ représentative de cette fonction, c'est-à-dire la nappe d'équation $z = f(x, y)$ est située, dans un voisinage du point $(a, f(a))$, au dessus (resp. dessous) de son plan tangent d'équation $z = f(a)$ et a l'allure de la nappe d'équation $z = x^2 + y^2$ (resp. $z = -x^2 - y^2$).

3. Une application : projection sur un convexe fermé

On donne une application à une situation particulière concernant la distance d'un point x à un sous-ensemble U relativement à la norme euclidienne $\|\cdot\|_2$. Pour $x \in \mathbb{R}^n$, on appelle *distance (euclidienne) de x à U* le nombre noté $d_U(x)$ et défini par

$$d_U(x) = \inf\{\|x-u\|_2;\ u \in U\}.$$

Il s'agit d'un problème idéal en optimisation ; on va montrer en effet l'existence, l'unicité et une caractérisation du point de U le plus proche de x.

Théorème 114. *Soit U un sous-ensemble convexe, fermé, non vide de $\mathbb{R}^n$.*

1 - Pour tout $x \in \mathbb{R}^n$, il existe un et un seul point de U, appelé projection de x sur U et noté $p_U(x)$, tel que

$$\|x - p_U(x)\|_2 = \inf\{\|x-u\|_2;\ u \in U\}.$$

En particulier $p_U(x) = x$ si $x \in U$.

2 - $p_U(x)$ est l'unique point de U vérifiant les inéquations $< x - p_U(x), u - p_U(x) > \leq 0$ pour tout $u \in U$.

3 - Pour tous x et $y \in \mathbb{R}^n$, on a $\|p_U(x) - p_U(y)\|_2 \leq \|x-y\|_2$.

Démonstration. 1 - (*Existence et unicité de $p_U(x)$*) L'existence de $p_U(x)$ est donnée par le théorème de Weierstrass appliqué à la fonction continue et cœrcive $f(u) = \|x-u\|_2^2$ et l'unicité de $p_U(x)$ est donnée par la proposition **??** car cette fonction est strictement convexe.

2 - (*Caractérisation de $p_U(x)$*) Soit $u \in U$ et $t \in]0,1]$. Comme $(1-t)p_U(x) + tu \in U$ on a

$$\begin{aligned} \|x-p_U(x)\|_2^2 &\leq \|x - ((1-t)p_U(x) + tu)\|_2^2 = \|x - p_U(x) - t(u - p_U(x))\|_2^2 \\ &= \|x-p_U(x)\|_2^2 - 2t < x - p_U(x), u - p_U(x) > + t^2\|u - p_U(x)\|_2^2. \end{aligned}$$

En simplifiant par $\|x - p_U(x)\|_2^2$, en divisant par t et en faisant tendre t vers 0 par valeurs > 0, on obtient $< x - p_U(x), u - p_U(x) > \leq 0$.

Inversement soit $z \in U$ tel que $< x - z, u - z > \leq 0$ pour tout $u \in U$. Alors

$$\|x-u\|_2^2 = \|x-z\|_2^2 + 2 < x - z, z - u > + \|z-u\|_2^2 \geq \|x-z\|_2^2$$

pour tout $u \in U$, d'où l'on déduit que $z = p_U(x)$ par unicité.

3 - (*p_U est 1-lipschitzienne*) En utilisant les inéquations précédentes caractérisant $p_U(x)$ et $p_U(y)$ on a

$$< x - p_U(x), p_U(y) - p_U(x) > \leq 0 \qquad < y - p_U(y), p_U(x) - p_U(y) > \leq 0.$$

En sommant ces inégalités et en utilisant l'inégalité de Cauchy-Schwarz, il vient

$$\|p_U(x) - p_U(y)\|_2^2 \leq < p_U(x) - p_U(y), x - y > \leq \|p_U(x) - p_U(y)\|_2 \|x-y\|_2,$$

d'où

$$\|p_U(x) - p_U(y)\|_2 \leq \|x-y\|_2.$$

Remarque. On peut donner d'autres démonstrations de l'existence et de l'unicité de la projection $p_U(x)$.

Pour l'existence de la projection $p_U(x)$, on peut montrer qu'une suite minimisante est convergente non pas en utilisant la propriété de Bolzano-Weierstrass de $\mathbb{R}^n$ (toute suite bornée de $\mathbb{R}^n$ contient une sous-suite convergente dans $\mathbb{R}^n$) comme on l'a fait pour le théorème de Weierstrass, mais la propriété de Cauchy de $\mathbb{R}^n$ (toute suite de Cauchy de $\mathbb{R}^n$ est convergente dans $\mathbb{R}^n$). Pour montrer qu'une suite minimisante $(x_k)_k$ est de Cauchy, on utilise l'identité du parallélogramme

$$\|z-y\|_2^2 + \|z+y\|_2^2 = 2(\|z\|_2^2 + \|y\|_2^2)$$

qu'on applique à $z = x - x_k$ et $y = x - x_h$.

Pour l'unicité de la projection $p_U(x)$, on peut aussi utiliser cette identité du parallélogramme qu'on applique à $z = x - p_1$ et $y = x - p_2$ si p_1 et p_2 sont deux points de U supposés réaliser le minimum $d_U(x)$.

Exemple. Si $U = \{x \in \mathbb{R}^n;\ \|x\|_2 \leq 1\}$ alors

$$\begin{aligned} p_U(x) = &\ x \quad && \text{si } \|x\|_2 \leq 1, \\ p_U(x) = &\ \frac{x}{\|x\|_2} && \text{si } \|x\|_2 > 1. \end{aligned}$$

En effet tout d'abord la boule unité U est convexe et fermée dans $\mathbb{R}^n$. Ensuite pour $\|x\|_2 > 1$ et $u \in U$, on a

$$\begin{aligned} < x - \frac{x}{\|x\|_2}, u - \frac{x}{\|x\|_2} > &= (\|x\|_2 - 1) < \frac{x}{\|x\|_2}, u - \frac{x}{\|x\|_2} > \\ &= \big(\|x\|_2 - 1\big)\big(< \frac{x}{\|x\|_2}, u > -1\big) \leq (\|x\|_2 - 1)(\|u\|_2 - 1) \leq 0. \end{aligned}$$

Ainsi on a bien $p_U(x) = \dfrac{x}{\|x\|_2}$ par la caractérisation de $p_U(x)$ du théorème **??**.

Exemple. Si $U = \{x \in \mathbb{R}^n;\ x_i \geq 0,\ i = 1, \ldots, n\}$, alors pour $x \in \mathbb{R}^n$

$$(p_U(x))_i = \sup(x_i, 0) \quad i = 1, \ldots, n.$$

(Ce résultat est suggéré en dimension $n = 2$ en examinant les différents cas de figures)

Tout d'abord l'ensemble U est convexe et fermé dans $\mathbb{R}^n$. Ensuite étant donné un élément quelconque $x = (x_1, \ldots, x_n)$ de $\mathbb{R}^n$ et posant $X = (X_1, \ldots, X_n)$ avec $X_i = \sup(x_i, 0)$ pour $i = 1, \ldots, n$, on a pour tout $u \in U$

$$< x - X, u - X >= \sum_{i=1}^{n} (x_i - X_i)(u_i - X_i) = \sum_{\substack{i=1 \\ x_i < 0}}^{n} x_i\, u_i \leq 0.$$

Ainsi on a bien $p_U(x) = X$ par la caractérisation de $p_U(x)$ du théorème **??**.

Un cas particulier important est celui où U est un sous-espace vectoriel de $\mathbb{R}^n$.

Corollaire 115. *(projection sur un sous-espace vectoriel) Soit U un sous-espace vectoriel de $\mathbb{R}^n$.*

1 - Pour tout $x \in \mathbb{R}^n$, il existe un et un seul point de U noté $p_U(x)$ tel que

$$\|x - p_U(x)\|_2 = \inf\{\|x - u\|_2;\ u \in U\}.$$

2 - $p_U(x)$ est l'unique point de U vérifiant les égalités $< x - p_U(x), u >= 0$ pour tout $u \in U$; on dit alors que $x - p_U(x)$ est orthogonal à U et que $p_U(x)$ est la projection orthogonale de x sur U.

3 - La fonction p_U est linéaire dans $\mathbb{R}^n$ avec $\|p_U(x)\|_2 \leq \|x\|_2$ pour tout $x \in \mathbb{R}^n$.

Démonstration. 1 - Un sous-espace vectoriel U de $\mathbb{R}^n$ étant convexe et fermé, on peut appliquer le théorème **??** de projection sur U.

2 - On sait que $p_U(x)$ vérifie $p_U(x) \in U$ et $< x - p_U(x), u - p_U(x) >\leq 0$ pour tout $u \in U$. En particulier en prenant $u = 2p_U(x)$ puis $u = 0$ qui appartiennent à U car U est un sous-espace vectoriel, on obtient d'abord $< x - p_U(x), p_U(x) >= 0$ et par suite

$$< x - p_U(x), u >\leq 0 \quad \text{pour} \quad \text{tout} \quad u \in U.$$

Puis en changeant $u \in U$ en $-u$ qui appartient aussi à U, on obtient bien que nécessairement

$$< x - p_U(x), u >= 0 \quad \text{pour} \quad \text{tout} \quad u \in U,$$

ce que l'on exprime en disant que $x - p_U(x)$ est orthogonal à U.

Inversement si $z \in U$ et vérifie $< x - z, u >= 0$ pour tout $u \in U$, alors $< x - z, u - z >= 0$ pour tout $u \in U$ car $u - z \in U$. Donc ce z vérifie la caractérisation de la projection $p_U(x)$.

3 - Soit x_1, $x_2 \in \mathbb{R}^n$ et $t \in \mathbb{R}$.

D'une part $p_U(x_1) + tp_U(x_2) \in U$ car U est un sous-espace vectoriel, et d'autre part pour $u \in U$, on a, par la bilinéarité du produit scalaire,

$$< x_1 + tx_2 - (p_U(x_1) + tp_U(x_2)), u >=< x_1 - p_U(x_1), u > +t < x_2 - p_U(x_2), u >$$

qui est nul par la caractérisation des projections $p_U(x_1)$ et $p_U(x_2)$ des points x_1 et x_2.

Donc $p_U(x_1) + tp_U(x_2) = p_U(x_1 + tx_2)$ par la caractérisation de la projection $p_U(x_1 + tx_2)$ du point $x_1 + tx_2$, ce qui montre la linéarité de p_U.

Enfin on a vu de façon générale que p_U est 1-lipschitzienne, donc dans la cas particulier présent elle vérifie $\|p_U(x)\|_2 \leq \|x\|_2$ car $0 \in U$ et $p_U(0) = 0$.

Exemple. Si U est un sous-espace vectoriel de $\mathbb{R}^n$ de dimension p et $f_1, \ldots, f_p$ une base orthonormée de U, c'est-à-dire une base de U telle que $< f_i, f_i >= 1$ pour $i = 1, \ldots, p$ et $< f_i, f_j >= 0$ si $i \neq j$, alors

$$p_U(x) = \sum_{i=1}^{p} < x, f_i > \ f_i.$$

Pour vérifier ce résultat, on utilise la caractérisation de $p_U(x)$ par les égalités d'orthogonalité $< x - p_U(x), u >= 0$ pour tout $u \in U$.

Chapitre 8 - Optimisation sous contraintes égalité

On s'intéresse dans ce chapitre au cas où l'ensemble des contraintes U est défini par des égalités de la forme

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \ldots, p\}$$

où Ω est un ouvert de $\mathbb{R}^n$ et $g_1, \ldots, g_p$ sont des fonctions définies dans Ω à valeurs dans $\mathbb{R}$.

En général, un tel ensemble U n'est pas ouvert, mais par contre il est souvent fermé (c'est le cas par exemple pour une courbe de $\mathbb{R}^2$ définie avec une fonction g continue). C'est pourquoi la condition nécessaire d'extremum libre donnée par l'équation d'Euler du théorème **??** du chapitre 6 ne s'applique certainement pas en général.

Dans ce cas on parle d'extremum *lié* par les contraintes $g_i(x) = 0$, $i = 1, \ldots, p$ qui sont aussi appelées les *liaisons*.

1. Optimisation sous contraintes égalité : condition nécessaire de Lagrange (d'ordre 1)

On donne d'abord la condition nécessaire suivante qui fait intervenir les dérivées premières de f par l'intermédiaire des multiplicateurs de Lagrange :

Théorème 116. *(condition nécessaire : multiplicateurs de Lagrange) Soit Ω un sous-ensemble ouvert non vide de $\mathbb{R}^n$, U le sous-ensemble de Ω défini par*

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \ldots, p\}$$

où $g_1,\ldots, g_p$ sont des fonctions définies dans Ω à valeurs dans $\mathbb{R}$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$ et a un point de U. Si

1 - a est un point d'extremum local de f dans U,

2 - f est différentiable en a,

3 - $g_1,\ldots, g_p$ sont de classe C^1 au voisinage de a,

4 - $\nabla g_1(a),\ldots,\nabla g_p(a)$ sont linéairement indépendants,

alors nécessairement il existe p réels $\lambda_1,\ldots,\lambda_p$, définis de façon unique, tels que

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \ldots + \lambda_p \nabla g_p(a).$$

Si les p vecteurs $\nabla g_1(a),\ldots,\nabla g_p(a)$ sont linéairement indépendants, c'est-à-dire si la matrice $(Jg(a))^T$, transposée de la matrice jacobienne de la fonction $g = (g_1, \ldots, g_p)$, est de rang p, on dit que les contraintes sont *qualifiées* au point a.

Les réels $\lambda_1,\ldots,\lambda_p$ tels que

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \ldots + \lambda_p \nabla g_p(a)$$

sont appelés les *multiplicateurs de Lagrange* associés au point d'extremum lié a et l'équation (vectorielle) $\nabla f(a) = \lambda_1 \nabla g_1(a)+\ldots+\lambda_p \nabla g_p(a)$ est appelée l'*équation de Lagrange* associée ;

elle correspond aux n équations (scalaires)

$$\frac{\partial f}{\partial x_i}(a) = \lambda_1 \frac{\partial g_1}{\partial x_i}(a) + ... + \lambda_p \frac{\partial g_p}{\partial x_i}(a), \qquad i = 1, \dots, n.$$

Démonstration. En supposant que les p vecteurs $\nabla g_1(a), \dots, \nabla g_p(a)$ sont linéairement indépendants (c'est-à-dire l'hypothèse 4 de qualification), on peut se ramener à une minimisation sans contrainte grâce au théorème des fonctions implicites.

Pour simplifier on fait la démonstration dans le cas $p = 1$ pour une seule contrainte notée g et on suppose que $\nabla g(a) \neq 0$, par exemple que $\dfrac{\partial g}{\partial x_n}(a) \neq 0$.

En notant $x \in \mathbb{R}^n$ sous la forme $x = (x', x_n)$ avec $x' = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$ et $x_n \in \mathbb{R}$, le théorème des fonctions implicites assure qu'il existe un voisinage ouvert V' de a' dans $\mathbb{R}^{n-1}$, un voisinage ouvert V_n de a_n dans $\mathbb{R}$ avec $V' \times V_n \subset \Omega$ et une fonction φ de V' dans V_n tels que

$$\begin{cases} (x', x_n) \in V' \times V_n \\ g(x', x_n) = 0 \end{cases} \Longleftrightarrow \begin{cases} x' \in V' \\ x_n = \varphi(x'). \end{cases}$$

De plus la fonction φ est de classe C^1 dans V' avec pour $i = 1, \dots, n-1$ et $x' \in V'$

$$\frac{\partial g}{\partial x_i}(x', \varphi(x')) + \frac{\partial g}{\partial x_n}(x', \varphi(x')) \, \frac{\partial \varphi}{\partial x_i}(x') = 0.$$

On définit alors la fonction F dans l'ouvert V' de $\mathbb{R}^{n-1}$ à valeurs dans $\mathbb{R}$ par

$$F(x') = f(x', \varphi(x')).$$

Cette fonction est différentiable en a' avec pour $i = 1, \dots, n-1$

$$\frac{\partial F}{\partial x_i}(a') = \frac{\partial f}{\partial x_i}(a) \, + \frac{\partial f}{\partial x_n}(a) \frac{\partial \varphi}{\partial x_i}(a').$$

De plus si a est un point de minimum relatif pour f dans U, alors le point a' est un point de minimum relatif pour F dans V'. En effet, quitte à remplacer V' et V_n par des ouverts plus petits, on a

$$f(a) \leq f(x) \quad \text{pour} \;\; x \in V' \times V_n.$$

Ainsi

$$f(a', \varphi(a')) \leq f(x', \varphi(x')) \quad \text{pour} \;\; x' \in V',$$

soit

$$F(a') \leq F(x') \quad \text{pour} \;\; x' \in V'.$$

Comme V' est un ouvert de $\mathbb{R}^{n-1}$, la condition nécessaire d'Euler assure que $\dfrac{\partial F}{\partial x_i}(a') = 0$ pour $i = 1, \dots, n-1$, soit d'après l'expression de ces dérivées partielles

$$\frac{\partial f}{\partial x_i}(a) = \frac{\partial f}{\partial x_n}(a) \left(\frac{\partial g}{\partial x_n}(a) \right)^{-1} \frac{\partial g}{\partial x_i}(a) \qquad \text{pour} \;\; i = 1, \dots, n-1.$$

Cette relation étant trivialement vraie pour $i = n$, en posant $\lambda = \dfrac{\partial f}{\partial x_n}(a)\left(\dfrac{\partial g}{\partial x_n}(a)\right)^{-1}$ on a donc la relation annoncée

$$\nabla f(a) = \lambda \, \nabla g(a)$$

avec un multiplicateur de Lagrange λ.

Dans le cas général on obtient p multiplicateurs de Lagrange et leur unicité est assurée par l'indépendance des gradients $\nabla g_1(a)$,...,$\nabla g_p(a)$ des contraintes. $\diamond$

Remarques. (sur l'hypothèse 4 de qualification des contraintes)

1 - Si $p = 1$, cette hypothèse de qualification signifie que $\nabla g_1(a) \neq 0$.

2 - Si l'hypothèse de qualification pour p contraintes est satisfaite, alors $p \leq n$, c'est-à-dire il y a moins de contraintes que de variables.

3 - Si l'hypothèse de qualification en un point d'extremum lié a n'est pas satisfaite, il se peut qu'il n'existe pas de multiplicateurs de Lagrange.

Ainsi par exemple soit $f(x, y) = x$ et la contrainte $g(x, y) = x^2 + y^2 = 0$. L'ensemble des contraintes U est réduit au point $(0, 0)$ et en ce point on a $\nabla g(0, 0) = (0, 0)$ et $\nabla f(0, 0) = (1, 0)$. Ainsi il n'existe pas de λ tel que $\nabla f(0, 0) = \lambda \nabla g(0, 0)$. Dans ce cas il n'existe donc pas de multiplicateur à cause de la dégénérescence de la contrainte au point $(0, 0)$.

4 - L'hypothèse de qualification en un point d'extremum lié a est une condition suffisante pour qu'il existe des multiplicateurs, mais elle n'est pas nécessaire.

Ainsi par exemple soit $f(x, y) = x^2 + y^2$ et les contraintes $g_1(x, y) = x^2 + y = 0$ et $g_2(x, y) = x^4 + y = 0$. Le point $(0, 0)$ de l'ensemble des contraintes U est le point de minimum global de f dans U. Les contraintes ne sont pas qualifiées en ce point car $\nabla g_1(0, 0) = \nabla g_2(0, 0) = (0, 1)$ et comme $\nabla f(0, 0) = (0, 0)$, on peut cependant écrire

$$\nabla f(0, 0) = \lambda \nabla g_1(0, 0) - \lambda \nabla g_2(0, 0).$$

5 - L'hypothèse de qualification porte sur le point de minimisation a qui est en pratique ce que l'on cherche et est donc a priori inconnu. Mais en général on aura toujours à considérer des ensembles de contraintes dont tous les points sont des points de qualification.

Autre formulation de l'équation de Lagrange.

Le *lagrangien* du problème de minimisation de f dans l'ensemble des contraintes $U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \dots, p\}$ est la fonction L définie dans $\Omega \times \mathbb{R}^p$ par

$$L(x, \lambda) = f(x) - < \lambda, g(x) >$$

c'est-à-dire

$$L(x, \lambda) = f(x) - \lambda_1 g_1(x) - \ldots - \lambda_p g_p(x)$$

si $\lambda = (\lambda_1, \dots, \lambda_p)$.

En particulier on a les trois identités

$L(x, \lambda) = f(x) \qquad$ pour $\ x \in U,\ \lambda \in \mathbb{R}^p$,

$$\frac{\partial L}{\partial x_i}(x,\lambda) = \frac{\partial f}{\partial x_i}(x,\lambda) - \lambda_1 \frac{\partial g_1}{\partial x_i}(x) - \ldots - \lambda_p \frac{\partial g_p}{\partial x_i}(x) \qquad \text{pour } x \in \Omega,\ \lambda \in \mathbb{R}^p, i = 1, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_j}(x,\lambda) = -g_j(x) \qquad \text{pour } x \in \Omega,\ \lambda \in \mathbb{R}^p, j = 1, \ldots, p.$$

Ainsi

$$\left\{ \begin{array}{l} (a,\lambda) \in \Omega \times \mathbb{R}^p \\ \nabla L(a,\lambda) = 0 \end{array} \right. \iff \left\{ \begin{array}{l} (a,\lambda) \in U \times \mathbb{R}^p \\ \nabla_x L(a,\lambda) = 0 \end{array} \right.$$

où $\nabla L(\cdot,\cdot)$ est le gradient de L dans les $n+p$ variables (x,λ) de $\Omega \times \mathbb{R}^p$ et $\nabla_x L(\cdot,\lambda)$ est le gradient de $L(\cdot,\lambda)$ dans les n variables x de Ω pour $\lambda \in \mathbb{R}^p$ fixé.

Par conséquent la condition nécessaire de Lagrange peut être exprimée de façon équivalente sous forme de point critique du Lagrangien L sous la forme suivante :

Théorème 116'. (*condition nécessaire : condition d'Euler pour le Lagrangien*)
Avec les notations et hypothèses du théorème **??**, *alors nécessairement il existe un $\lambda \in \mathbb{R}^p$ défini de façon unique, tel que (a,λ) soit un point critique de L dans $\Omega \times \mathbb{R}^p$, c'est-à-dire vérifie l'équation d'Euler*

$$\nabla L(a,\lambda) = 0.$$

Mise en œuvre pratique du théorème 116'. Pratiquement on cherchera d'abord les points de qualification a des contraintes, puis parmi ces points a on cherchera à mettre en évidence les éventuels points d'extremum a de f dans U en cherchant les points critiques (a,λ) du lagrangien L dans $\Omega \times \mathbb{R}^p$ par la résolution des équations d'Euler-Lagrange : on écrit que les $n+p$ inconnues a_i pour $i = 1, \ldots, n$ et λ_j pour $j = 1, \ldots, p$ sont solutions du système de $n+p$ équations

$$\left\{ \begin{array}{rcl} \dfrac{\partial f}{\partial x_1}(a) - \lambda_1 \dfrac{\partial g_1}{\partial x_1}(a) - \ldots - \lambda_p \dfrac{\partial g_p}{\partial x_1}(a) & = & 0 \\ & \vdots & \\ \dfrac{\partial f}{\partial x_n}(a) - \lambda_1 \dfrac{\partial g_1}{\partial x_n}(a) - \ldots - \lambda_p \dfrac{\partial g_p}{\partial x_n}(a) & = & 0 \\ g_1(a) & = & 0 \\ & \vdots & \\ g_p(a) & = & 0. \end{array} \right.$$

On notera que c'est la transposée $\big(Jg(a)\big)^T$ de la matrice jacobienne de la fonction $g = (g_1, \ldots, g_p)$ qui intervient et non la matrice jacobienne $Jg(a)$ elle-même.

La contrepartie de la prise en compte de la contrainte $g(x) = (g_1(x), \ldots, g_p(x)) = 0$ est donc la résolution d'un système plus gros que celui associé sans contrainte. En effet on ne peut pas se dispenser de calculer le vecteur $\lambda = (\lambda_1, \ldots, \lambda_p)$ de $\mathbb{R}^p$ même si, comme c'est fréquemment le cas, on ne s'intéressera qu'aux seuls extrema éventuels a de f dans U, l'inconnue λ apparaissant alors simplement comme un intermédiaire nécessaire de calcul.

Il ne faut pas perdre de vue que les conditions ci-dessus ne sont que des conditions nécessaires sur a pour que a soit un extremum de f dans U. Si l'on veut affirmer que les points a mis en évidence par ces conditions nécessaires correspondent bien à un maximum ou un minimum relatif a de la fonction f dans l'ensemble des contraintes U, et si l'on ne peut pas appliquer les conditions suffisantes que l'on verra plus loin, il faudra compléter ces conditions nécessaires par une étude locale.

Exemple. (contraintes égalité) On cherche les extrema de $f(x) = x_1 + \ldots + x_n$ dans $\mathbb{R}^n$ sous la contrainte $x_1^2 + \ldots + x_n^2 = 1$, autrement dit on cherche les points d'extremum de la fonction f dans le sous-ensemble $U = \{x \in \mathbb{R}^n;\ g(x) = 0\}$ où g est la fonction définie dans $\mathbb{R}^n$ par $g(x) = x_1^2 + \ldots + x_n^2 - 1$.

Tout d'abord comme f est continue dans $\mathbb{R}^n$ et que la sphère unité U est fermée bornée, il existe au moins un point qui est un point de minimum global de f dans U et un point qui est un point de maximum global de f dans U.

Comme $\nabla g(x) = 2x$, ce vecteur est non nul pour $x \in U$. Par conséquent tous les points de U sont des points de qualification de la contrainte et les conditions du théorème de Lagrange sont satisfaites.

On cherche donc les éventuels points d'extremum x de f dans U en cherchant les points critiques (x, λ) du lagrangien $L(x, \lambda) = f(x) - \lambda g(x)$ dans $\mathbb{R}^n \times \mathbb{R}$ par la résolution des équations d'Euler pour L dans $\mathbb{R}^n \times \mathbb{R}$

$$\left\{\begin{array}{rcl} 1 - 2\lambda x_1 & = & 0 \\ & \vdots & \\ 1 - 2\lambda x_n & = & 0 \\ x_1^2 + \ldots + x_n^2 - 1 & = & 0. \end{array}\right.$$

Ces équations d'Euler ont pour solutions

$$\Big(x = -\frac{1}{\sqrt{n}}(1, \ldots, 1), \lambda = -\frac{\sqrt{n}}{2}\Big) \qquad \text{et} \qquad \Big(x = +\frac{1}{\sqrt{n}}(1, \ldots, 1), \lambda = +\frac{\sqrt{n}}{2}\Big).$$

Ainsi nécessairement, si a est un point d'extremum de f dans U, alors $a = -\dfrac{1}{\sqrt{n}}(1, \ldots, 1)$ ou $a = +\dfrac{1}{\sqrt{n}}(1, \ldots, 1)$.

Or $f\big(-\dfrac{1}{\sqrt{n}}(1, \ldots, 1)\big) = -\sqrt{n}$ et $f\big(+\dfrac{1}{\sqrt{n}}(1, \ldots, 1)\big) = +\sqrt{n}$ et comme il existe au moins un point qui est un point de minimum global de f dans U et un point qui est un point de maximum global de f dans U, on en déduit que

- le point $-\dfrac{1}{\sqrt{n}}(1, \ldots, 1)$ est l'unique point de minimum global strict de $x_1 + \ldots + x_n$ sur la sphère $x_1^2 + \ldots + x_n^2 = 1$,

- le point $+\dfrac{1}{\sqrt{n}}(1,\ldots,1)$ est l'unique point de maximum global strict de $x_1+\ldots+x_n$ sur la sphère $x_1^2+\ldots+x_n^2=1$. $\diamond$

Exemple. (contraintes inégalité) On cherche les extrema de la fonction f définie dans $\mathbb{R}^2$ par $f(x,y)=xy$ dans le sous-ensemble $U=\{(x,y)\in\mathbb{R}^2;\ x^2+y^2-2\le 0\}$.

Tout d'abord comme f est continue dans $\mathbb{R}^2$ et que la boule fermée U (de centre $(0,0)$ et de rayon $\sqrt{2}$ pour la norme euclidienne) est fermée bornée, il existe au moins un point qui est un point de minimum global de f dans U et un point qui est un point de maximum global de f dans U.

On examine les deux possibilités d'existence : dans l'intérieur $\overset{\circ}{U}=\{(x,y)\in\mathbb{R}^2;\ g(x,y)<0\}$ et sur le bord $\partial U=\{(x,y)\in\mathbb{R}^2; g(x,y)=0\}$ avec $g(x,y)=\ x^2+y^2-2$.

1 - Dans l'intérieur $\overset{\circ}{U}=\{(x,y)\in\mathbb{R}^2;\ g(x,y)<0\}$, le seul point critique de f est le point $(0,0)$. Or ce point n'est pas un point d'extremum de f, c'est un point col car par exemple $f(x,-x)\,(=-x^2)<f(0,0)<f(x,x)\,(=x^2)$ pour $x\in\mathbb{R}$, $x\neq 0$.

Ainsi il n'y a pas de point d'extremum de f dans $\overset{\circ}{U}$.

2 - Sur le bord $\partial U=\{(x,y)\in\mathbb{R}^2;\ g(x,y)=0\}$, on a $\nabla g(x,y)=(2x,2y)$ et ce vecteur est non nul pour $(x,y)\in\partial U$. Par conséquent tous les points de ∂U sont des points de qualification de la contrainte g et les conditions du théorème de Lagrange sont satisfaites.

On cherche donc les éventuels points d'extremum de f dans ∂U en cherchant les points critiques $((x,y),\lambda)$ du lagrangien $L(x,y,\lambda)=xy-\lambda(x^2+y^2-2)$ dans $\mathbb{R}^2\times\mathbb{R}$ par la résolution des équations d'Euler pour L dans $\mathbb{R}^2\times\mathbb{R}$

$$\left\{\begin{array}{rcll} y-2\,\lambda\,x & = & 0 & \qquad (1)\\ x-2\,\lambda\,y & = & 0 & \qquad (2)\\ x^2+y^2-2 & = & 0 & \qquad (3). \end{array}\right.$$

En reportant l'équation (1) dans (2), on doit avoir nécessairement

$$x(1-4\,\lambda^2)=0$$

d'où

- soit $x=0$ et par suite $y=0$ d'après (1), ce qui est impossible par (3),
- soit $\lambda=\pm\dfrac{1}{2}$ et par suite $y=\pm x$ d'après (1) et $y=\pm 1$ d'après (3).

Ainsi les point critiques $((x,y),\lambda)$ du lagrangien $L(x,y,\lambda)$ dans $\mathbb{R}^2\times\mathbb{R}$ sont

$$\big((1,1),\frac{1}{2}\big)\quad\big((-1,-1),\frac{1}{2}\big)\quad\big((1,-1),-\frac{1}{2}\big)\quad\big((-1,1),-\frac{1}{2}\big).$$

Par suite nécessairement les points d'extremum de f dans ∂U sont parmi les points $(1,1)$, $(-1,-1)$, $(-1,1)$ et $(1,-1)$.

Or $f(1,1)=f(-1,-1)=1$ et $f(-1,1)=f(1,-1)=-1$, et comme il existe au moins un point qui est un point de minimum global de f dans ∂U et un point qui est un point de maximum global de f dans ∂U, on en déduit que

- les points $(1,1)$ et $(-1,-1)$ sont les points de maximum global de f dans la boule U,
- les points $(-1,1)$ et $(1,-1)$ sont les points de minimum global de f dans la boule U.

Ces extrema ne sont pas stricts car ils sont atteints en deux points distincts. $\diamond$

Exemple. On cherche les extrema de la fonction f définie dans $\mathbb{R}^2$ par $f(x,y) = -y$ dans le sous-ensemble $U = \{(x,y) \in \mathbb{R}^2;\ g(x,y) = 0\}$ avec $g(x,y) = x^2 + y^2 - 1$.

Tout d'abord comme f est continue dans $\mathbb{R}^2$ et que la sphère unité U (de centre $(0,0)$ et de rayon 1 pour la norme euclidienne) est fermée bornée, il existe au moins un point qui est un point de minimum global de f dans U et un point qui est un point de maximum global de f dans U.

Comme $\nabla g(x,y) = (2x, 2y)$ est non nul pour $(x,y) \in U$, tous les points de U sont des points de qualification de la contrainte g et les conditions du théorème de Lagrange sont satisfaites.

On cherche donc les éventuels points d'extremum (x,y) de f dans U en cherchant les points critiques $((x,y),\lambda)$ du lagrangien $L(x,y,\lambda) = -y - \lambda(x^2+y^2-1)$ dans $\mathbb{R}^2 \times \mathbb{R}$ par la résolution des équations d'Euler pour L dans $\mathbb{R}^2 \times \mathbb{R}$

$$\left\{\begin{array}{rcll} -2\,\lambda\,x & = & 0 & \qquad (1) \\ -1-2\,\lambda\,y & = & 0 & \qquad (2) \\ x^2+y^2-1 & = & 0 & \qquad (3). \end{array}\right.$$

Par l'équation (1), on doit avoir nécessairement
- soit $\lambda = 0$ et par suite $-1 = 0$ d'après (2), ce qui est impossible,
- soit $x = 0$ et par suite $y = \pm 1$ d'après (3) et $\lambda = \mp\dfrac{1}{2}$ d'après (2).

Ainsi les point critiques $((x,y),\lambda)$ du lagrangien $L(x,y,\lambda)$ dans $\mathbb{R}^2 \times \mathbb{R}$ sont

$$\Big((0,1), -\frac{1}{2}\Big) \qquad \Big((0,-1), +\frac{1}{2}\Big).$$

Par suite nécessairement les points d'extremum de f dans U sont parmi les points $(0,1)$ et $(0,-1)$.

Or $f(0,1) = -1$ et $f(0,-1) = +1$ et comme il existe au moins un point qui est un point de minimum global de f dans U et un point qui est un point de maximum global de f dans U, on en déduit que
- le point $(0,1)$ est le point de maximum global strict de f sur la sphère U,
- le point $(0,-1)$ est le point de minimum global strict de f sur la sphère U. $\diamond$

2. Optimisation sous contraintes égalité : condition suffisante de Lagrange

Si (a,λ) est un point critique du lagrangien L dans l'ouvert $\Omega \times \mathbb{R}^p$, alors on a vu par la condition nécessaire de Lagrange (si a est un point de qualification des contraintes) que a est un point candidat pour être un extremum lié de la fonction f dans l'ensemble des contraintes U.

On va chercher des conditions suffisantes pour que a soit effectivement un extremum de f dans U.

On note de plus que a est aussi un point critique de $L(\cdot,\lambda)$ dans l'ouvert Ω qui appartient à U, et donc par la condition nécessaire d'Euler a est un point de U candidat pour être un extremum libre de $L(\cdot,\lambda)$ dans l'ouvert Ω.

En fait on va montrer dans le théorème qui suit que si un point a de U est un extremum (libre) de la fonction $L(\cdot,\lambda)$ dans l'ouvert Ω alors il est un extremum (lié) de la fonction f dans l'ensemble des contraintes U.

Théorème 117. *Soit Ω un sous-ensemble ouvert non vide de $\mathbb{R}^n$, U le sous-ensemble de Ω défini par*

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \dots, p\}$$

où g_1,...,g_p sont des fonctions définies dans Ω à valeurs dans $\mathbb{R}$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$ et L le lagrangien associé.

Si un point a de U est un point d'extremum (libre) de $L(\cdot,\lambda)$ dans l'ouvert Ω pour un λ de $\mathbb{R}^p$, alors a est un point d'extremum (lié) de même nature (global, local, minimum, maximum) de f dans l'ensemble des contraintes U.

Si f, $g_1, \dots, g_p$ sont différentiables en a, on note que (a,λ) est alors nécessairement un point critique de L dans $\Omega \times \mathbb{R}^p$ et donc que $\lambda = (\lambda_1, \dots, \lambda_p)$ est un multiplicateur de Lagrange associé au point a d'extremum (lié) de f dans U.

Démonstration. Supposons par exemple que a soit un point de minimum global de $L(\cdot,\lambda)$ dans Ω, c'est-à-dire

$$L(a,\lambda) \leq L(x,\lambda) \qquad x \in \Omega;$$

en particulier

$$L(a,\lambda) \leq L(x,\lambda) \qquad x \in U.$$

Or $f(x) = L(x,\lambda)$ pour $x \in U$. Par conséquent si de plus a est un point de U, alors

$$f(a) \leq f(x) \qquad x \in U.$$

Ainsi a est un point de minimum global de f dans U. ◇

Ce théorème montre l'intérêt de la notion de Lagrangien ; parmi les points critiques (a,λ) de L dans $\Omega \times \mathbb{R}^p$ elle ramène la recherche d'extremum lié de la fonction f dans U à la recherche d'extremum libre de la fonction $L(\cdot,\lambda)$ dans l'ouvert Ω.

Remarque. Cette proposition est une condition suffisante mais non nécessaire de façon générale pour avoir un extremum de f dans U ; si (a,λ) est un point critique de L dans $\Omega \times \mathbb{R}^p$, il se peut que a soit un extremum de f sous la contrainte U sans que a soit un extremum de $L(\cdot,\lambda)$ dans Ω.

Suite au théorème **??** on peut alors écrire deux situations dans lesquelles la condition nécessaire de Lagrange pour un extremum lié sous contraintes d'égalité est en fait une condition suffisante.

On donne tout d'abord une condition suffisante de convexité sur le Lagrangien.

Corollaire 118. *(condition suffisante : convexité de $L(\cdot,\lambda)$) Soit Ω un sous-ensemble ouvert, convexe, non vide de $\mathbb{R}^n$, U le sous-ensemble de Ω défini par*

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \dots, p\}$$

où g_1,...,g_p sont des fonctions dans Ω à valeurs dans $\mathbb{R}$, f une fonction dans Ω à valeurs dans $\mathbb{R}$ et L le lagrangien associé.

Si (a,λ) est un point critique de L dans l'ouvert $\Omega \times \mathbb{R}^p$ (en supposant f, g_1,...,g_p différentiables en a) et si $L(\cdot,\lambda)$ est convexe (resp. concave) dans l'ouvert Ω alors a est un point de minimum (resp. maximum) global de f dans U.

Démonstration. Comme Ω est un ouvert convexe de $\mathbb{R}^n$ et que $L(\cdot,\lambda)$ est convexe dans Ω, alors a est minimum global de $L(\cdot,\lambda)$ dans Ω si et seulement si a est un point critique de $L(\cdot,\lambda)$, c'est-à-dire $\nabla_x L(a,\lambda) = 0$ (d'après le corollaire 4 du chapitre 7). Or si (a,λ) est un point critique de L dans $\Omega \times \mathbb{R}^p$ alors en particulier $a \in U$ et $\nabla_x L(a,\lambda) = 0$, donc a appartient à U et est un point de minimum global de $L(\cdot,\lambda)$ dans Ω. Le théorème **??** permet alors de conclure. $\diamond$

En particulier si la fonction coût f est convexe et les contraintes g_i sont affines, on a :

Corollaire 119. *Soit Ω un sous-ensemble ouvert, convexe, non vide de $\mathbb{R}^n$, U le sous-ensemble de Ω défini par*

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \dots, p\}$$

où g_1,...,g_p sont des fonctions affines dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, f une fonction convexe (resp. concave) dans Ω à valeurs dans $\mathbb{R}$ et L le lagrangien associé.

Si (a,λ) est un point critique de L dans l'ouvert $\Omega \times \mathbb{R}^p$ (en supposant f différentiable en a) alors a est un point de minimum (resp. maximum) global de f dans U.

Enfin on donne une condition suffisante de positivité de la matrice hessienne en x du lagrangien.

Corollaire 120. *(condition suffisante : positivité de $H_x L(\cdot,\lambda)$) Soit Ω un sous-ensemble ouvert, non vide de $\mathbb{R}^n$, U le sous-ensemble de Ω défini par*

$$U = \{x \in \Omega;\ g_i(x) = 0,\ i = 1, \dots, p\},$$

où g_1,...,g_p sont des fonctions définies dans Ω à valeurs dans $\mathbb{R}$, f une fonction définie dans Ω à valeurs dans $\mathbb{R}$ et L le lagrangien associé.

Si (a,λ) est un point critique de L dans l'ouvert $\Omega \times \mathbb{R}^p$ et si la matrice hessienne de $L(\cdot,\lambda)$ en a est définie positive (resp. négative) (en supposant $f, g_1, \dots, g_p$ deux fois différentiables en a), alors a est un point de minimum (resp. maximum) local strict de f dans U.

Démonstration. Par hypothèse la fonction $L(\cdot,\lambda) = f - \sum_{i=1}^{p} \lambda_i\, g_i$, définie dans l'ouvert Ω, est deux fois différentiable en a, le point a est un point critique dans Ω, c'est-à-dire $\nabla_x L(a,\lambda) = 0$ et sa matrice hessienne $H_x L(a,\lambda)$ est définie positive.

Par conséquent par la condition forte de Legendre de la proposition **??** du chapitre 6, ce point a est un point de maximum local strict de $L(\cdot,\lambda)$ dans Ω. Comme de plus $a \in U$ car $\nabla L_\lambda(a,\lambda) = 0$, le théorème **??** permet d'en déduire que ce point a est un point de maximum local strict de f dans U. $\diamond$

Chapitre 9 - Optimisation quadratique

Dans ce chapitre on applique les résultats généraux vus précédemment à quelques problèmes de programmation quadratique, c'est-à-dire à des problèmes d'optimisation dans lesquels la fonction f à optimiser est une fonction quadratique dans $\mathbb{R}^n$, autrement dit un polynôme de degré inférieur ou égal à 2 des variables $x_1, \ldots, x_n$.

1. Un problème d'optimisation sans contraintes : résolution d'un système linéaire

Un exemple d'optimisation sans contrainte dans $\mathbb{R}^n$ provient de la résolution des systèmes linéaires dans $\mathbb{R}^n$. Plus précisément on va montrer que la résolution d'un système linéaire du type $Ax = b$ dans $\mathbb{R}^n$ équivaut à la recherche d'un minimum de la fonction quadratique $f(x) = \dfrac{1}{2} < Ax, x > - < b, x >$ dans $\mathbb{R}^n$.

Ce résultat théorique a des conséquences pratiques importantes pour la résolution des systèmes linéaires : on cherche à résoudre le problème de minimisation par des algorithmes numériques pour obtenir les solutions du système linéaire correspondant.

Théorème 121. *Soit f une fonction quadratique dans $\mathbb{R}^n$ de la forme*

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

où A est une matrice symétrique semi-définie positive de taille (n, n) et b un vecteur de $\mathbb{R}^n$. Alors

1 - il existe au moins un point de $\mathbb{R}^n$ qui est un point de minimum global de f dans $\mathbb{R}^n$,

2 - ces points de minimum sont les points critiques de f dans $\mathbb{R}^n$, c'est-à-dire les solutions du système linéaire

$$Ax = b,$$

3 - il existe un et un seul point de $\mathbb{R}^n$ qui est un point de minimum global strict de f dans $\mathbb{R}^n$, si et seulement si la matrice A est définie positive.

Démonstration. On peut vérifier ce résultat en notant par exemple que f est coercive, différentiable et convexe (strictement si A est définie positive) dans $\mathbb{R}^n$ avec $\nabla f(x) = Ax - b$, et en appliquant les résultats des chapitres précédents, en particulier le corollaire **??** du chapitre 7. $\diamond$

2. Un problème d'optimisation sans contraintes : approximation au sens des moindres carrés

Un autre exemple d'optimisation sans contraintes dans $\mathbb{R}^n$ provient d'un problème d'approximation. Etant donnés p points distincts x_i de $\mathbb{R}$ pour $i = 1, \ldots, p$, p valeurs numériques associées c_i pour $i = 1, \ldots, p$ et un ensemble de n fonctions w_j pour $j = 1, \ldots, n$ linéairement indépendantes, définies sur un ensemble contenant les p points x_i, on cherche à déterminer une fonction u de la forme

$$u(x) = a_1 w_1(x) + \ldots + a_n w_n(x)$$

telle que les p égalités $u(x_i) = c_i$ pour $i = 1, \ldots, p$, soient approchées au mieux.

La façon la plus commune de réaliser cette approximation consiste à approcher ces p égalités $u(x_i) = c_i$ au sens des moindres carrés, c'est-à-dire on cherche une fonction $u(x) = a_1 w_1(x) + \ldots + a_n w_n(x)$ qui rend minimum le nombre

$$\sum_{i=1}^{p} |\sum_{j=1}^{n} a_j w_j(x_i) - c_i|^2$$

lorsque le vecteur $a = (a_j)_{1 \leq j \leq n}$ décrit $\mathbb{R}^n$.

Le théorème suivant donne une réponse à ce problème :

Théorème 122. *Soit f une fonction définie dans $\mathbb{R}^n$ de la forme*

$$f(x) = \|Bx - c\|_{2p}^2$$

où B est une matrice de taille (p, n) avec $p \geq n$, c un vecteur de $\mathbb{R}^p$ et $\| \cdot \|_{2p}$ la norme euclidienne dans $\mathbb{R}^p$. Alors

1 - il existe au moins un point de $\mathbb{R}^n$ qui est un point de minimum global de f dans $\mathbb{R}^n$,

2 - ces points de minimum sont les solutions du système linéaire

$$B^T B x = B^T c,$$

3 - il existe un et un seul point de $\mathbb{R}^n$ qui est un point de minimum global strict de f dans $\mathbb{R}^n$ si et seulement si la matrice B est de rang n.

Les équations linéaires $B^T B x = B^T c$ sont appelées les équations normales associées au problème d'approximation au sens des moindres carrés.

Démonstration 1. (par application du théorème précédent) On introduit la fonctionnelle quadratique F définie dans $\mathbb{R}^n$ par

$$F(x) = \frac{1}{2}\|Bx - c\|_{2p}^2 - \frac{1}{2}\|c\|_{2p}^2$$

qui est égale à

$$F(x) = \frac{1}{2} < B^T B x, x >_n - < B^T c, x >_n$$

où $< \cdot, \cdot >_n$ désigne le produit scalaire dans $\mathbb{R}^n$.

La matrice $B^T B$ est symétrique et semi-définie positive car $< B^T B x, x >_n = \|Bx\|_{2p}^2$. De plus elle est définie positive si et seulement si B est injective, c'est-à-dire si et seulement si B est de rang n.

Le théorème **??** permet alors de conclure. $\diamond$

Démonstration 2. (par application du théorème de projection) Le théorème de projection sur le sous-espace vectoriel

$$Im\,B = \{Bx \in \mathbb{R}^p;\ x \in \mathbb{R}^n\}$$

entraîne l'existence et l'unicité d'un élément $\tilde{a}$ tel que

$$\tilde{a} \in Im\,B \quad \text{et} \quad \|\tilde{a} - c\|_{2p} = \inf\{\|\tilde{x} - c\|_{2p};\ \tilde{x} \in Im\,B\}.$$

Par suite le problème de minimisation a comme solutions les points $a \in \mathbb{R}^n$ tels que $Ba = \tilde{a}$.

Or la projection $\tilde{a}$ est caractérisée par

$$< \tilde{a} - c, \tilde{x} >_p = 0 \quad \text{pour} \quad \tilde{x} \in Im\,B$$

donc les points a de minimisation sont caractérisés par

$$< Ba - c, Bx >_p = 0 \quad \text{pour} \quad x \in \mathbb{R}^n.$$

Comme $< Ba - c, Bx >_p = < B^T Ba - B^T c, x >_n$, on retrouve ainsi qu'un point a de $\mathbb{R}^n$ est solution du problème de minimisation si et seulement si il est solution du système linéaire

$$B^T Ba = B^T c$$

et que ces équations normales ont toujours au moins une solution. ◇

Remarque. Par la formule de Taylor à l'ordre 2 on a pour tous x et $h \in \mathbb{R}^n$

$$F(x+h) = F(x) + < \nabla F(x), h >_n + \frac{1}{2} < B^T Bh, h >_n$$

soit

$$\|B(x+h) - c\|_{2p}^2 = \|Bx - c\|_{2p}^2 + 2 < B^T Bx - B^T c, h >_n + \|Bh\|_{2p}^2.$$

On retrouve ainsi qu'un point x de $\mathbb{R}^n$ est solution du problème de minimisation si et seulement si il est solution des équations normales.

Remarque. 1 - Par abus de langage on dit parfois que les points de minimisation trouvés ci-dessus sont les solutions du système linéaire $Bx = c$ au sens des moindres carrés.

2 - Si $p = n$ et si B est inversible, la solution des équations normales coïncide avec la solution du système linéaire $Bu = c$.

3 - Si B est de rang n, l'application qui à c associe la solution du système linéaire $Bx = c$ au sens des moindres carrés, est linéaire.

3. Un problème d'optimisation avec contraintes affines

Un exemple d'optimisation avec contraintes de type affine provient de la discrétisation par éléments finis des équations de Stokes ; ces équations sont un modèle simplifié d'écoulement d'un fluide visqueux incompressible, les vraies équations étant celles de Navier-Stokes.

Théorème 123. *Soit f une fonction quadratique dans $\mathbb{R}^n$ de la forme*

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

où A est une matrice symétrique définie positive de taille (n, n) et b un vecteur de $\mathbb{R}^n$, et soit U l'ensemble des contraintes

$$U = \{x \in \mathbb{R}^n;\, Cx = d\}$$

où C est une matrice de taille (p, n) de rang p et d un vecteur de $\mathbb{R}^p$.

Alors il existe un et un seul point de $\mathbb{R}^n$ qui est un point de minimum global strict de f dans U.

Démonstration. Posant $g(x) = Cx - d$, on a $Jg(x) = C$ et par conséquent chaque point x de $\mathbb{R}^n$ est un point de qualification des contraintes g. On cherche donc les éventuels points d'extremum de f dans U en cherchant les points critiques (x, λ) du lagrangien $L(x, \lambda) = f(x) - < \lambda,\, g(x) >_p$ dans $\mathbb{R}^n \times \mathbb{R}^p$ par la résolution des équations d'Euler de L dans $\mathbb{R}^n \times \mathbb{R}^p$

$$\nabla L(x, \lambda) = 0$$

soit

$$\begin{cases} Ax - C^T\lambda = b & (1) \\ \quad\quad\; Cx = d & (2). \end{cases}$$

Or ce système linéaire de $n + p$ équations à $n + p$ inconnues a une et une seule solution si et seulement si le système homogène associé

$$\begin{cases} Ax - C^T\lambda = 0 & (1') \\ \quad\quad\; Cx = 0 & (2') \end{cases}$$

a la seule solution $x = 0$, $\lambda = 0$. Or de (1') on a

$$< Ax, x > - < C^T\lambda, x >= 0$$

soit

$$< Ax, x > - < \lambda, Cx >= 0$$

et avec (2') on a

$$< Ax, x >= 0.$$

Comme A est définie positive, on en déduit que $x = 0$, et revenant à (1') on a

$$C^T\lambda = 0.$$

Comme C est de rang p, on en déduit que $\lambda = 0$.

Ainsi il existe un et un seul point critique (a, λ) du lagrangien L dans $\mathbb{R}^n \times \mathbb{R}^p$. Par conséquent comme la fonction f est convexe dans $\mathbb{R}^n$ et les composantes g_i de g sont affines, par le corollaire **??** du chapitre 8 par exemple, le point a est un point de minimum global de f dans U. De plus par la condition nécessaire de Lagrange (puisque tous les points de U sont de qualification) il est le seul point d'extremum de f dans U. $\diamond$

4. Un problème d'optimisation avec contraintes dans la sphère unité : valeurs propres et quotients de Rayleigh

On donne la caractérisation de la plus petite et de la plus grande valeurs propres d'une matrice symétrique.

Théorème 124. *Soit A une matrice symétrique de taille (n,n) et λ_i pour $i=1,\ldots,n$ ses n valeurs propres (qui sont réelles). Alors*

$$\begin{aligned}\inf_{1\leq i\leq n}\lambda_i &= \inf\left\{\frac{<Ax,x>}{\|x\|_2^2};\, x\in\mathbb{R}^n,\, x\neq 0\right\}\\ \sup_{1\leq i\leq n}\lambda_i &= \sup\left\{\frac{<Ax,x>}{\|x\|_2^2};\, x\in\mathbb{R}^n,\, x\neq 0\right\}.\end{aligned}$$

Démonstration. On note que

$$\left\{\frac{<Ax,x>}{\|x\|_2^2};\, x\in\mathbb{R}^n,\, x\neq 0\right\}=\{<Ax,x>;\, x\in\mathbb{R}^n,\, \|x\|_2=1\},$$

les quotients $\dfrac{<Ax,x>}{\|x\|_2^2}$ étant appelés les quotients de Rayleigh de la matrice A.

On cherche à minimiser la fonction quadratique f définie dans $\mathbb{R}^n$ par $f(x)=<Ax,x>$ dans la sphère unité $U=\{\,x\in\mathbb{R}^n,\, \|x\|_2=1\}$.

La fonction f est continue de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et U est fermé et borné dans $\mathbb{R}^n$. Donc par le théorème de Weierstrass, il existe au moins un point a de minimum global de f dans U.

Posant $g(x)=\|x\|_2^2$, cette fonction g est différentiable dans $\mathbb{R}^n$ avec $\nabla g(x)=2x$ qui est non nul dans U, donc tous les points de U sont des points de qualification de la contrainte. De plus la fonction f est différentiable dans $\mathbb{R}^n$ avec $\nabla f(x)=A$, donc par la condition nécessaire de Lagrange, il existe un multiplicateur que l'on note $\dfrac{1}{2}\lambda$ dans $\mathbb{R}$ tel que

$$Aa=\lambda a.$$

Donc λ est une valeur propre de A associée au vecteur propre a de $\mathbb{R}^n$ normalisé par $\|a\|_2=1$. En particulier

$$\lambda=\lambda<a,a>=<Aa,a>=\inf\{<Ax,x>;\, x\in\mathbb{R}^n,\, \|x\|_2=1\}.$$

Si ν est une autre valeur propre de A associée à un vecteur propre b de $\mathbb{R}^n$ normalisé par $\|b\|_2=1$, on a

$$\nu=\nu<b,b>=<Ab,b>\geq\inf\{<Ax,x>;\, x\in\mathbb{R}^n,\, \|x\|_2=1\}.$$

Donc λ est la plus petite valeur propre de A, c'est-à-dire

$$\inf_{1\leq i\leq n}\lambda_i=\inf\{<Ax,x>;\, x\in\mathbb{R}^n,\, \|x\|_2=1\}.$$

On montre de même la formule du sup en maximisant la fonction f dans U. $\diamond$

Exercices

TD 1. L'espace $\mathbb{R}^n$

Exercice 1.

Déterminer et représenter graphiquement les ensembles $\{(x,y)\in\mathbb{R}^2, x^2+y^2-2x-2y+1=0\}$, $\{(x,y)\in\mathbb{R}^2, (16-x^2-y^2)(x^2+y^2-4)>0\}$, $\{(x,y,z)\in\mathbb{R}^3, 2x+4y+3z=12\}$ et enfin $\{(x,y,z)\in\mathbb{R}^3, 3x^2+6x+3y^2+3z^2=0\}$.

Exercice 2.

Pour $x=(x_1,\ldots,x_n)$ dans $\mathbb{R}^n$ on pose

$$\|x\|_1=\sum_{i=1}^n |x_i|,\quad \|x\|_2=\Big(\sum_{i=1}^n x_i^2\Big)^{\frac{1}{2}},\quad \|x\|_\infty=\sup_{i=1,\ldots,n}|x_i|.$$

a) Si $<x,y>$ désigne le produit scalaire euclidien des deux vecteurs x et y de $\mathbb{R}^n$, montrer l'inégalité de Cauchy-Schwarz :

$$|<x,y>|\le \|x\|_2\,\|y\|_2.$$

b) Montrer que les applications $\|\cdot\|_p$ sont des normes sur $\mathbb{R}^n$ pour $p=1,\ 2,\ \infty$ et que pour tout x dans $\mathbb{R}^n$ on a

$$\frac{1}{\sqrt{n}}\|x\|_1\le\|x\|_2\le\|x\|_1.$$
$$\|x\|_\infty\le\|x\|_1\le n\,\|x\|_\infty,$$
$$\|x\|_\infty\le\|x\|_2\le\sqrt{n}\,\|x\|_\infty\,.$$

c) En déduire que ces 3 normes sont équivalentes.

d) Dans le cas où $n=2$, déterminer et représenter graphiquement le cercle de centre 0 et de rayon 1 pour chacune de ces 3 normes.

Exercice 3.

On se place dans l'espace euclidien $\mathbb{R}^n$.

a) Soit x dans $\mathbb{R}^n$ tel que $<x,y>=0$ pour tout y dans $\mathbb{R}^n$. Montrer que $x=0$.

b) On dit qu'un ensemble A de $\mathbb{R}^n$ est dense si pour tout y dans $\mathbb{R}^n$ il existe une suite d'éléments de A convergeant vers y. Soit alors A un tel ensemble et x dans $\mathbb{R}^n$ tels que $<x,y>=0$ pour tout y dans A. Montrer que $x=0$.

c) Montrer que $\|x\|_2=\sup\limits_{y\in\mathbb{R}^n,\|y\|=1}<x,y>=\sup\limits_{y\in\mathbb{R}^n,\|y\|\le1}<x,y>$.

d) Montrer l'identité du parallélogramme :

$$\|x+y\|_2^2+\|x-y\|_2^2=2\,(\|x\|_2^2+\|y\|_2^2).$$

Exercice 4.

Soit $(u_n)_n$ une suite réelle convergente, de limite u. Montrer que l'ensemble $\{u_n, n\in\mathbb{N}\}\cup\{u\}$ est un fermé borné de $\mathbb{R}$.

Exercice 5.

Soit A un ensemble de $\mathbb{R}^n$ muni de la norme $\|.\|$. Pour x dans $\mathbb{R}^n$ on note $d(x, A) = \inf_{a \in A} \|x - a\|$ la distance de x à A.

a) Montrer que $|d(x, A) - d(y, A)| \leq \|x - y\|$ pour tous x et y dans $\mathbb{R}^n$.

b) Si A est fermé et $x \notin A$, montrer que $d(x, A) > 0$.

c) Si A est fermé borné, montrer qu'il existe a dans A tel que $d(x, A) = \|x - a\|$.

d) Dans le cas où $n = 1$ et $A = \{x \in \mathbb{Q},\, x^2 < 2\}$, calculer $d(2, A)$. Existe-t-il a dans A tel que $d(2, A) = \|2 - a\|$?

Exercice 6.

Soit A et B deux ensembles de $\mathbb{R}^n$ muni de la norme $\|.\|$, et notons $D(A, B) = \inf_{a \in A, b \in B} \|a - b\|$ la distance de A à B.

a) Montrer que $D(A, B) = \inf_{a \in A} d(a, B) = \inf_{b \in B} d(b, A)$ avec la notation de l'exercice précédent.

b) Si A est fermé borné, montrer qu'il existe a dans A tel que $D(A, B) = d(a, B)$.

c) Si A et B sont fermés bornés, montrer qu'il existe $a \in A$ et $b \in B$ tels que $D(A, B) = \|a - b\|$.

d) Si A et B sont disjoints, avec A fermé borné et B fermé, montrer que $D(A, B) > 0$.

e) Donner des exemples d'ensembles A et B de $\mathbb{R}^2$ muni de la norme $\|(x, y)\| = |x| + |y|$, tels que

i) A et B soient fermés (mais pas bornés) et $D(A, B) = 0$;

ii) A et B ne soient pas fermés et $D(A, B) = 0$;

iii) A soit fermé borné, B non fermé et $D(A, B) = 0$.

Exercice 7.

Si $B(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$ désigne la boule ouverte de centre x et rayon $r > 0$ dans $\mathbb{R}^n$, on note $\overset{o}{A} = \{x \in \mathbb{R}^n, \exists r > 0, B(x, r) \subset A\}$ l'intérieur d'un ensemble A de $\mathbb{R}^n$ et $\overline{A} = \{x \in \mathbb{R}^n, \forall r > 0, B(x, r) \cap A \neq \emptyset\}$ son adhérence. Soit A et B deux ensembles de $\mathbb{R}^n$.

a) Montrer que $\overset{o}{A}$ est le plus grand ouvert contenu dans A.

b) Montrer que $\overline{A}$ est le plus petit fermé contenant A, et que x appartient à $\overline{A}$ si et seulement si x est limite d'une suite de points de A.

c) Si $A \subset B$, montrer que $\overset{o}{A} \subset \overset{o}{B}$ et $\overline{A} \subset \overline{B}$.

d) Montrer que $A = \overline{A}$ si et seulement si A est fermé, que $A = \overset{o}{A}$ si et seulement si A est ouvert, et que $\overset{o}{A}$ est vide si et seulement si tout point de A est limite d'une suite de points du complémentaire de A.

e) Montrer que l'adhérence de $A \cup B$ est $\overline{A} \cup \overline{B}$ et que l'intérieur de $A \cap B$ est $\overset{o}{A} \cap \overset{o}{B}$.

f) Soit $A =]0, 2[\cup]3, 4[$ et $B =]1, 3[\cup]4, 5[$. Montrer que les ensembles $A \cap \overline{B}$, $\overline{A} \cap B$, $\overline{A} \cap \overline{B}$ et $\overline{A \cap B}$ sont deux à deux distincts.

TD 2. Continuité

Exercice 1.
Etudier la continuité de l'application $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x,y) = (x^2+y^2)^a \sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) \text{ pour } (x,y) \neq (0,0) \text{ et } f(0,0) = 0\,.$$

Exercice 2.
Les fonctions suivantes sont-elles continues sur $\mathbb{R}^2$?

$$f_1(x,y) = \frac{(x+y)^2}{x^2+y^2}\,, \qquad f_2(x,y) = \frac{x^3+y^3}{x^2+y^2}\,, \qquad f_3(x,y) = -4xy\frac{x^2-y^2}{x^2+y^2}$$

$$f_4(x,y) = \frac{|x+y|}{\sqrt{x^2+y^2}}\,, \qquad f_5(x,y) = \frac{x - sinx}{x^2+y^2}\,, \qquad f_6(x,y) = \frac{xy}{x^2+y^4}$$

pour $(x,y) \neq (0,0)$, et $f_i(0,0) = 0$ pour $i = 1, ..., 6$.

Exercice 3.
On définit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ par $f(x,y) = \dfrac{xy^2}{x^2+y^4}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$.
1) Montrer que la restriction de f à toute droite passant par l'origine est continue en $(0,0)$.
2) Montrer que f n'est pas continue en $(0,0)$.

Exercice 4.
Soit f la fonction définie sur $\{(x,y) \in \mathbb{R}^2 ; x \neq y\}$ par $f(x,y) = \dfrac{\sin x - \sin y}{x-y}$.
Montrer que l'on peut définir $f(x,x)$ pour $x \in \mathbb{R}$ de sorte que la fonction f ainsi prolongée soit continue sur $\mathbb{R}^2$.

Exercice 5.
Soit une application $f : \mathbb{R}^n \to \mathbb{R}^p$.
1) Montrer que si f est continue, alors pour tout ensemble ouvert A de $\mathbb{R}^p$, l'ensemble $f^{-1}(A) = \{x \in \mathbb{R}^n, f(x) \in A\}$ est un ouvert de $\mathbb{R}^n$.
2) Etudier la réciproque.
3) Le résultat est il encore vrai si on remplace ouvert par fermé ?
4) Soit A et B deux fermés non vides et disjoints de $\mathbb{R}^n$. Si $\|.\|$ désigne une norme sur $\mathbb{R}^n$, on pose $f(x) = d(x,A) - d(x,B)$ où $d(x,E) = \inf\{\|x-y\|, y \in E\}$. Montrer que f est continue et en déduire qu'il existe deux ouverts disjoints U et V tels que $A \subset U$ et $B \subset V$.

Exercice 6.

Soit f et g deux applications de $\mathbb{R}$ dans $\mathbb{R}$ et h l'application de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par $h(x,y) = f(x) + g(y)$ pour tout $(x,y) \in \mathbb{R}^2$.

1) Montrer que si f continue en x_0 et si g est continue en y_0, alors h est continue en (x_0, y_0).

2) On suppose h continue en (x_0, y_0). L'application f est elle continue en en x_0 ? L'application g est elle continue en en y_0 ?

Exercice 7.

Montrer que les ensembles

$$E = \{(a,b,c) \in \mathbb{R}^3; a \neq 0\}$$

et

$$\Omega = \{a,b,c) \in \mathbb{R}^3; a \neq 0, \text{le trinôme } ax^2 + bx + c \text{ a deux racines réelles distinctes}\}$$

sont deux ouverts de $\mathbb{R}^3$.

Exercice 8.

Si f est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$, montrer que l'ensemble $\{(x,y) \in \mathbb{R}^2; y = f(x)\}$ est un fermé de $\mathbb{R}^2$.

Exercice 9.

Montrer que l'ensemble $\{(x_1, \cdots, x_n) \in \mathbb{R}^n; x_i \geq 0 \, \forall i, \displaystyle\sum_{i=1}^{n} x_i = 1\}$ est un fermé borné de $\mathbb{R}^n$.

Exercice 10.

Si $a = (a_1, \cdots, a_p)$ est un vecteur de $\mathbb{R}^p$, déterminer la norme de l'application linéaire $f : \mathbb{R} \to \mathbb{R}^p$ définie par $f(t) = (a_1 t, \cdots, a_p t)$ lorsque $\mathbb{R}^p$ est muni des normes $\|\cdot\|_1, \|\cdot\|_2$ puis $\|\cdot\|_\infty$.

Exercice 11.

Si $a = (a_1, \cdots, a_n)$ est un vecteur de $\mathbb{R}^n$, déterminer la norme de l'application linéaire $f : \mathbb{R}^n \to \mathbb{R}$ définie par $f(x_1, \cdots, x_n) = \displaystyle\sum_{i=1}^{n} a_i x_i$ lorsque $\mathbb{R}^n$ est muni des normes $\|\cdot\|_1, \|\cdot\|_2$ puis $\|\cdot\|_\infty$.

Exercice 12.

Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ une application linéaire de matrice $A = (a_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}}$ dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^p$. Déterminer sa norme quand $\mathbb{R}^n$ et $\mathbb{R}^p$ sont munis de la même norme $\|\cdot\|_1$ puis de la même norme $\|\cdot\|_\infty$.

TD 3 - Différentiabilité

Exercice 1.
Montrer que la fonction $f : \mathbb{R} \to \mathbb{R}^3, t \mapsto (e^t, \sin t, \cos t)$ est dérivable sur $\mathbb{R}$ et calculer sa dérivée.

Exercice 2.
Montrer que le domaine de définition des fonctions suivantes est un ouvert et que ces fonctions sont de classe C^1 sur leurs domaines de définitions. Calculer leurs fonctions dérivées partielles premières :

$$a)\ f(x,y) = x^\alpha y^\beta \quad (\alpha > 0, \beta > 0) \qquad b)\ g(x,y) = x^y$$

$$c)\ h(x,y) = \exp(xy)\ln(1+x^2+y^2) \qquad d)\ k(x,y) = \frac{y\sin(xy)}{\sqrt{1-x^2-y^2}}$$

Exercice 3.
Soit $\Gamma = \big\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\ ;\ \forall i \in \{1, 2, \ldots, n\}, x_i > 0\big\}$.

(1) Montrer que Γ est ouvert.

(2) On considère les fonctions suivantes où f est supposée définie sur Γ :

$$f(x_1, x_2, \ldots, x_n) = x_1 + x_2^{1/2} + \cdots + x_i^{1/i} + \cdots + x_n^{1/n}, \qquad g(x_1, x_2, \ldots, x_n) = x_1 x_2^2 x_3^3 \ldots x_i^i \ldots x_n^n.$$

Montrer que ces fonctions sont de classe C^1 sur leurs domaines de définition et calculer leurs dérivées partielles premières.

Exercice 4.
On considère l'application f définie par :

$$f(x,y,z) = \big(\ln\sqrt{x^2+y^2+z^2}, xy\exp(z), x^2+y^3+4z\big).$$

(1) Déterminer le domaine de définition de f et montrer que c'est un ouvert de $\mathbb{R}^3$.

(2) Montrer que f est de classe C^1 sur son domaine de définition. Calculer la matrice jacobienne $Jf(M)$ pour $M = (x,y,z)$.

(3) Soit $M_0 = (1,1,0)$. Calculer la norme de l'application linéaire $df(M_0)$ quand $\mathbb{R}^3$ est muni de la norme $\|.\|_\infty$, puis de la norme $\|.\|_1$.

Exercice 5.
Soit $g : \mathbb{R} \to \mathbb{R}$ une application de classe C^2 et $f : \mathbb{R}^2 \to \mathbb{R}$ dfinie par $f(x,y) = \dfrac{g(x)-g(y)}{x-y}$ si $x \neq y$ et $f(x,x) = g'(x)$. Montrer que f est de classe C^1 sur $\mathbb{R}^2$ et calculer sa différentielle.

Exercice 6.
Soit $f(x,y) = (x^2+y^4)^{1/3}$.

(1) Montrer que f est définie sur $\mathbb{R}^2$ et de classe C^1 sur $\mathbb{R}^2\setminus\{(0,0)\}$. Calculer les dérivées partielles premières de f en un point $(x,y) \neq (0,0)$.

(2) La fonction f admet-elle des d'erivées partielles au point $(0,0)$?

Exercice 7.
Soit $f(x,y) = \dfrac{xy^2}{x^2+y^2}$ pour $(x,y) \neq (0,0)$ et $f(0,0) = 0$.

(1) Montrer que f est continue en $(0,0)$.

(2) Calculer $\dfrac{\partial f}{\partial x}(0,0)$ et $\dfrac{\partial f}{\partial y}(0,0)$.

(3) Montrer que f n'est pas différentiable en $(0,0)$.

Exercice 8.
Soit la fonction f définie par $f(x,y) = (x^2+y^2)\sin\Big(\dfrac{1}{\sqrt{x^2+y^2}}\Big)$ pour $(x,y) \neq (0,0)$ et $f(0,0) = 0$. Montrer que f est différentiable en $(0,0)$ mais pas de classe C^1 en $(0,0)$.

Exercice 9. Soit la fonction g définie par $g(x,y) = \dfrac{x^2y^2}{x^2+y^2}$ pour $(x,y) \neq (0,0)$ et $g(0,0) = 0$.

(1) Montrer que g est continue sur $\mathbb{R}^2$.

(2) Calculer les dérivées partielles premières de g au point $(0,0)$.

(3) Déterminer les fonctions dérivées partielles premières de g. Montrer que g est de classe C^1 sur $\mathbb{R}^2$.

Exercice 10.
On munit $\mathbb{R}^n$ de la norme $\|.\|_2$ associée au produit scalaire euclidien.

(1) Soit g la fonction définie sur $\mathbb{R}^n$ par $g(x) = \|x\|_2^2 + 1$. Montrer que g est différentiable sur $\mathbb{R}^n$ et calculer $\nabla g(x)$ ainsi que $dg(x)(h)$ pour x et h dans $\mathbb{R}^n$.

(2) Soit f la fonction définie sur $\mathbb{R}^n$ par $f(x) = \dfrac{1}{\|x\|_2^2+1} = \dfrac{1}{g(x)}\cdot$ Montrer que f est différentiable sur $\mathbb{R}^n$ et calculer $\nabla f(x)$ pour $x \in \mathbb{R}^n$. Montrer que $df(x)(h) = -\dfrac{2 < x,h >}{(1+\|x\|_2^2)^2}$ pour tout x,h dans $\mathbb{R}^n$.

(3) Calculer $\|\nabla f(x)\|_2$ en fonction de $\|x\|_2$ pour tout $x \in \mathbb{R}^n$ et montrer que $\|\nabla f(x)\|_2 \leq 1$.

TD 4. Dérivées partielles, développements de Taylor et fonctions implicites

Exercice 1.

Soient α un réel et $D = \{x \in \mathbb{R}^n; x_i > 0\, \forall i = 1, \ldots, n\}$. Une fonction f de classe C^1 sur D est dite homogène de degré α si $f(tx) = t^\alpha f(x)$ pour tous $t > 0$ et $x \in D$.

a) Vérifier que D est un ouvert de $\mathbb{R}^n$ et que $tx \in D$ pour tout $t > 0$ et $x \in D$.

b) Montrer que la fonction f définie sur D par $f(x) = x_1 x_2^2 \ldots x_n^n$ est homogène et déterminer son degré.

c) Soit f une fonction homogène de degré α sur D, $x \in D$ fixé, et g la fonction définie sur $\mathbb{R}_*^+$ par $g(t) = f(tx)$.

Calculer la dérivée de g de deux manières différentes et en déduire que $< x, \nabla f(x) >= \alpha f(x)$.

Supposons que $\alpha = 1$ et soit $h = rx$ où r est un réel strictement positif. Exprimer $f(x+h)$ à l'aide du développement limité d'ordre 1 de f au voisinage de x et montrer que le reste est nul.

Exercice 2.

Soit f la fonction définie sur $\mathbb{R}^2$ par $f(x,y) = xy\dfrac{x^2 - y^2}{x^2 + y^2}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$.
Montrer que les dérivées partielles secondes $\dfrac{\partial^2 f}{\partial x \partial y}(0,0)$ et $\dfrac{\partial^2 f}{\partial y \partial x}(0,0)$ existent et calculer leur valeur.

Quelle conclusion peut-on en tirer ?

Exercice 3.

Soit g, h et $k : \mathbb{R}^2 \to \mathbb{R}$ les fonctions définies par

$$g(x,y) = x + y, \quad h(x,y) = x + y^2, \quad k(x,y) = x^2 + y.$$

a) Dans chacun des cas suivants déterminer, si elle existe, une fonction $f \in C^2(\mathbb{R}^2)$ telle que

$$\frac{\partial f}{\partial x} = g,\ \frac{\partial f}{\partial y} = h; \qquad \frac{\partial f}{\partial x} = g,\ \frac{\partial f}{\partial y} = k; \qquad \frac{\partial f}{\partial x} = h,\ \frac{\partial f}{\partial y} = k.$$

b) Discuter, dans les cas d'existence, l'eventuelle unicité : si l'on trouve deux telles fonctions f_1 et f_2, que peut-on dire de $f_1 - f_2$?

Exercice 4.
Soit A la matrice suivante :

$$A = \begin{pmatrix} 0 & 2 \\ -1 & 2 \end{pmatrix}.$$

a) Donner un exemple, s'il existe, de fonction $f : \mathbb{R}^2 \to \mathbb{R}^2$ de matrice jacobienne A en 0.
b) Donner un exemple, s'il existe, de fonction $f \in C^2(\mathbb{R}^2)$ de matrice hessienne A en 0.
c) Donner un exemple, s'il existe, de fonction $f \in C^2(\mathbb{R}^2)$ de matrice hessienne $A + A^T$ en 0.

Exercice 5.
Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x, y) = \sin(x + y) - \cos(x - y) - e^{x+y} + 2$.
a) Montrer que f est de classe C^2 sur $\mathbb{R}^2$ et déterminer le plus petit entier p tel qu'il existe une dérivée partielle d'ordre p de f en $(0, 0)$ non nulle.
b) Écrire le développement de Taylor de f en $(0, 0)$ jusqu'à l'ordre p inclus.
c) Déterminer la limite éventuelle

$$\lim_{(x,y)\to(0,0)} \frac{f(x, y)}{(\sqrt{x^2 + y^2})^p}.$$

Exercice 6.
Écrire le développement de Taylor d'ordre 2 en $(0, 0)$ de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x, y) = \cos\left(\tan(x + \sin y)\right).$$

Exercice 7.
Soit $f \in C^2(\mathbb{R}^n)$ une fonction admettant une constante M telle que

$$\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right| \leq M \qquad \text{pour tout } 1 \leq i, j \leq n,\ x \in \mathbb{R}^n.$$

a) Montrer qu'il existe des constantes A, B et C telles que $|f(x)| \leq A + B\|x\| + C\|x\|^2$ pour tout $x \in \mathbb{R}^n$. La réponse dépend-t-elle de la norme $\|\cdot\|$ choisie ?
b) Peut-on estimer la constante C en termes de M ? et les constantes A et B ?
c) Montrer qu'il existe des constantes D et E telles que $|f(x)| \leq D + E\|x\|^2$ pour tout $x \in \mathbb{R}^n$. Est-il possible d'estimer E en termes de M comme précédemment ?

Exercice 8.
a) Montrer que la courbe Γ de $\mathbb{R}^2$ d'équation $x^3 + y^3 - 3xy = 1$ est, localement autour du point $(0, 1)$, d'équation $y = \varphi(x)$. Déterminer le développement limité d'ordre 2 de φ au voisinage de 0 et l'équation de la tangente à Γ en $(0, 1)$.
b) Montrer que la surface Σ de $\mathbb{R}^3$ d'équation $x^2+y^2+z^2+2xyz = 1$ est, localement autour du point $(0, 0, 1)$, d'équation $z = \psi(x, y)$. Déterminer le développement limité d'ordre 1 de ψ au voisinage de $(0, 0)$ et l'équation du plan tangent à Σ en $(0, 0, 1)$.

TD 5. Convexité

Exercice 1.

Soit U un sous-ensemble convexe de $\mathbb{R}^n$. Une fonction f sur U à valeurs dans $\mathbb{R}$ est dite affine si $f(tx+(1-t)y) = tf(x) + (1-t)\,f(y)$ pour $x,\, y \in U$ et $t \in [0,1]$.

a) Si a est un vecteur de $\mathbb{R}^n$ et b un réel, montrer que l'application f définie sur $\mathbb{R}^n$ par $f(x) =< a, x > +b$ est affine.

b) Soit f une fonction affine sur U et g une fonction convexe sur un intervalle I de $\mathbb{R}$ tel que $f(U) \subset I$. Montrer que la fonction $g \circ f$ est convexe sur U.

c) En déduire que la fonction h définie par $h(x,y) = \ln(1+2x+3y)$ est concave sur son ensemble de définition, que l'on déterminera.

Exercice 2.

Montrer qu'une norme sur $\mathbb{R}^n$ est convexe et que sa restriction à la demi-droite $d_+(0;u) = \{tu; t \geq 0\}$, pour $u \in \mathbb{R}^n$ fixé, est une fonction affine ; en déduire qu'elle n'est pas strictement convexe sur $\mathbb{R}^n$.

Exercice 3.

Soit U un convexe non vide de $\mathbb{R}^n$ muni de la norme $\|.\|$. Pour $x \in \mathbb{R}^n$, on note $d_U(x) = \inf\{\|x-u\|;\ u \in U\}$ la distance de x à U. Montrer que les fonctions d_U et d_U^2 sont convexes sur $\mathbb{R}^n$.

Exercice 4.

Soit f une fonction convexe sur un convexe U de $\mathbb{R}^n$ et $E_t = \{x \in U; f(x) \leq t\}$ son ensemble de niveau $t \in \mathbb{R}$. Montrer que E_t est un convexe de $\mathbb{R}^n$.

Exercice 5.

Montrer que les fonctions f et g définies sur $\mathbb{R}^2$ par $f(x,y) = x^4+y^4$ et $g(x,y) = (x-y)^2$ sont convexes sur $\mathbb{R}^2$, mais que $h = f - g$ n'est ni convexe ni concave sur $\mathbb{R}^2$.75

Exercice 6.

a) Quel est le domaine de définition D de la fonction f définie par $f(x,y) = \dfrac{1-xy}{x+y}$? D est-il ouvert ? fermé ? borné ? convexe ?

b) Etudier la convexité de f sur $D_1 = \{(x,y) \in \mathbb{R}^2; x+y > 0\}$ et $D_2 = \{(x,y) \in \mathbb{R}^2; x+y < 0\}$.

c) Etudier la convexité de $g(x,y) = \dfrac{1-xy}{x+y} - \sqrt{x+y}$ sur son domaine de définition.

TD 6. Optimisation sans contrainte

Exercice 1.
Déterminer les extrema locaux des fonctions suivantes sur $\mathbb{R}^2$:

$$\begin{array}{rclrcl} f_1(x,y) &=& x^3+3xy^2-15x-12y, & f_2(x,y) &=& 3x^3+xy^2-xy, \\ f_3(x,y) &=& x^4+\frac{1}{3}y^3-4y-2, & f_4(x,y) &=& x^3+xy^2-x^2y-y^3. \end{array}$$

Pour chaque fonction, montrer que les extrema locaux ne sont pas globaux.

Exercice 2.
Déterminer les extrema locaux des fonctions suivantes sur leur domaine de définition :

$$f(x,y,z) = x + \frac{y^2}{4x} + \frac{z^2}{y} + \frac{2}{z}$$

$$g(x,y,z) = 8x^3 + y^3 - 12xyz + 10z^3 - 6z.$$

Exercice 3.
Optimiser les fonctions suivantes sur leur domaine de définition :

$$\begin{array}{rcl} f(x,y,z) &=& x^4+2y^2+3z^2-yz-23y+4x-5 \\ g(x_1,\dots,x_n) &=& \displaystyle\sum_{i=1}^n x_i \ln\left(\frac{1}{x_i}\right) \\ h(x_1,\dots,x_n) &=& \displaystyle\prod_{i=1}^n x_i^{x_i} \text{ (on pourra utiliser la fonction ln)} \end{array}$$

Exercice 4.
Soit f la fonction définie sur $\mathbb{R}^2$ par $f(x,y) = x^4 + y^4 - (x-y)^2$.
a) Déterminer les extremums locaux de f sur $\mathbb{R}^2$.
b) Montrer que f admet un minimum global sur $\mathbb{R}^2$ (on pourra montrer que f est coercive).

Exercice 5.
Soit f la fonction définie sur $\mathbb{R}^3$ par $f(x,y,z) = x^4 - 2x^2y + 2y^2 - 2yz + 2z^2 - 4z + 5$.
a) Déterminer les points critiques de f sur $\mathbb{R}^3$.
b) Montrer que l'expression $f(x,y,z)$ peut s'écrire sous forme d'une somme de carrés.
c) En déduire les extrema de f sur $\mathbb{R}^3$.

Exercice 6. Droite des moindres carrés.

Soit n un entier, $n \geq 2$. On considère n points (x_i, y_i) de $\mathbb{R}^2$, pour $i = 1, \dots, n$ et la fonction f définie sur $\mathbb{R}^2$ par

$$f(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

a) Vérifier que f n'admet qu'un seul point critique $(\hat{a}, \hat{b})$ sur $\mathbb{R}^2$. Exprimer $\hat{b}$ en fonction de $\hat{a}$.

b) Montrer que f admet un minimum global sur $\mathbb{R}^2$.

Exercice 7.

Soient f et g les fonctions définies par $f(x, y) = xy$ et $g(x, y) = \dfrac{1}{x} + \dfrac{1}{y} \cdot$ En se ramenant à l'optimisation de fonctions d'une variable, déterminer

a) les extrema de f sur $\mathbb{R}^2$ puis sous la contrainte $g(x, y) = \dfrac{2}{3}$;

b) les extrema de g sur son domaine de définition puis sous la contrainte $f(x, y) = 9$.

TD 7. Optimisation sous contrainte

Exercice 1.

Déterminer les extrema locaux et globaux de la fonction f sous la contrainte $g = 0$ avec

$$\begin{array}{llllll} \text{a)} & f(x,y) & = & -y\,, & g(x,y) & = & x^2 + y^2 - 1\,; \\ \text{b)} & f(x,y) & = & x + y\,, & g(x,y) & = & \exp\left(x^2 + y^2 - \frac{1}{4}\right) - 1\,. \end{array}$$

Exercice 2.

Déterminer les extrema locaux et globaux de la fonction f sous la contrainte $g = 0$ avec

$$\begin{array}{llllll} \text{a)} & f(x,y) & = & x + 2y\,, & g(x,y) & = & x^2 + xy + y^2 + y - \dfrac{13}{9}\,; \\ \text{b)} & f(x,y) & = & x^2 + y^2\,, & g(x,y) & = & 4x^2 - y^2 - 16\,; \\ \text{c)} & f(x,y) & = & \ln(x - y)\,, & g(x,y) & = & x^2 + y^2 - 2\,; \end{array}$$

Exercice 3.

Reprendre l'exercice 7 de la feuille de TD 6 en utilisant la notion de lagrangien.

Exercice 4.
Déterminer les extrema locaux et globaux de la fonction f sous la contrainte $g = 0$ avec

$$\begin{array}{llcllcl} \text{a)} & f(x,y) & = & xy\,, & g(x,y) & = & x^2+y^2-x-y\,; \\ \text{b)} & f(x,y) & = & \dfrac{1}{x}+\dfrac{1}{y}\,, & g(x,y) & = & \dfrac{1}{x^2}+\dfrac{1}{y^2}-\dfrac{1}{2}\,; \\ \text{c)} & f(x,y) & = & x^3+y^3\,, & g(x,y) & = & x^2+y^2-1\,. \end{array}$$

Exercice 5.
Déterminer les extrema locaux et globaux de la fonction f sous la contrainte $g = 0$ avec

$$\begin{array}{llcllcl} \text{a)} & f(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i \ln x_i\,, & g(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i - 1\,; \\ \text{b)} & f(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i^2\,, & g(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i - 1\,; \\ \text{c)} & f(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i\,, & g(x_1,\dots,x_n) & = & \displaystyle\sum_{i=1}^n x_i^2 - 1\,. \end{array}$$

Donner le lien entre l'inégalité de Cauchy-Schwarz et le résultat des questions b) et c).

Exercice 6.
a) Soit E le sous-ensemble $\{(x,y) \in \mathbb{R}^2; x \geq 0, y \geq 0, x+y \leq 1\}$ de $\mathbb{R}^2$.
Déterminer les extrema globaux sur E de la fonction f définie par

$$f(x,y) = x^2 - 2y^2 - 5xy.$$

b) Même question avec $f(x,y) = 4x^2 + 5y^2 + 6xy - 6x - 6y + 3$.
c) Même question avec $E = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 - 2 \leq 0\}$ et $f(x,y) = xy$.

Sujets d'examen

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Contrôle continu du 26 mars 2007

Questions de cours

1. Donner la définition d'un ensemble ouvert, d'un ensemble fermé et d'un ensemble borné de $\mathbb{R}^n$.

2. Montrer que, si O est un ouvert de $\mathbb{R}^p$ et f une fonction continue de $\mathbb{R}^n$ dans $\mathbb{R}^p$, alors l'ensemble $f^{-1}(O) = \{x \in \mathbb{R}^n; f(x) \in O\}$ est un ensemble ouvert de $\mathbb{R}^n$.

Exercice

1. Pour tous réels x et y, montrer que $x^2 - x\,y + y^2 \geq \dfrac{1}{2}\big(x^2 + y^2\big)$.

2. Etant donné le réel a on note E_a l'ensemble défini par

$$E_a = \{(x, y) \in \mathbb{R}^2; x^2 - x\,y + y^2 = a^2\}.$$

a) Montrer que E_a est un ensemble fermé de $\mathbb{R}^2$.

b) Montrer que E_a est un ensemble borné de $\mathbb{R}^2$. Pour cela on pourra utiliser le résultat de la question 1.

c) Déterminer l'ensemble E_0.

3. Etant donnés deux réels strictement positifs p et q, on considère la fonction f définie sur $\mathbb{R}^2$ par

$$f(x, y) = \begin{cases} \dfrac{x^p\,y^q}{x^2 - x\,y + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

a) Montrer que la fonction f est continue sur $\mathbb{R}^2 \setminus \{(0, 0)\}$.

b) Montrer que les deux applications partielles de f en $(0, 0)$ sont continues en 0.

c) Sous quelle condition sur p et q la restriction de f à $\{(x, y) \in \mathbb{R}^2; x = y\}$ est-elle continue au point $(0, 0)$?

d) Sous quelle condition sur p et q la fonction f est-elle continue en $(0, 0)$? On pourra utiliser le résultat de la question 1.

4. a) Montrer que la fonction f est minorée et majorée sur l'ensemble E_1, et qu'elle y atteint ses bornes.

b) On suppose $p = q = 1$. Montrer que, pour tout $(x, y) \in E_1$, $f(x, y) = xy = x^2 + y^2 - 1$.
En déduire que, pour tout $(x, y) \in E_1$, $-1 \leq f(x, y) \leq 1$.
Le maximum de f sur E_1 est-il égal à 1 ?
Le minimum de f sur E_1 est-il égal à -1 ?

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Examen partiel du 5 avril 2007

Les documents, calculatrices et téléphones portables sont interdits pendant la durée de l'épreuve.

Question de cours
Soit $\mathbb{R}^n$ muni de la norme $\|.\|$, et f une fonction continue de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que $f(x)$ tende vers $+\infty$ quand $\|x\|$ tend vers $+\infty$ (la fonction f est dite coercive). Montrer que f est minorée et atteint son minimum.

Exercice 1
Soit f la fonction à valeurs réelles définie sur $\mathbb{R}^2$ par

$$f(x,y) = \begin{cases} \dfrac{\sin^4 x}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

1. a) Etablir l'inégalité $|\sin x| \leq |x|$ pour tout réel x.
b) Montrer que la fonction f est continue sur $\mathbb{R}^2$.

2. a) Montrer que f est de classe C^1 sur $\mathbb{R}^2 \setminus \{(0,0)\}$ et calculer ses dérivées partielles premières.
b) Montrer que f admet des dérivées partielles premières au point $(0,0)$, et les calculer.
c) Montrer que pour tous x et y

$$\left|\frac{\partial f}{\partial x}(x,y)\right| \leq 6|x| \qquad \text{et} \qquad \left|\frac{\partial f}{\partial y}(x,y)\right| \leq 2|y|.$$

En déduire que f est de classe C^1 sur $\mathbb{R}^2$.

Exercice 2
Soit g la fonction de $\mathbb{R}^3$ dans $\mathbb{R}^3$ définie par

$$g(x,y,z) = (x+y^2, y+z^2, z+x^2).$$

1. Montrer que g est de classe C^1 sur $\mathbb{R}^3$ et calculer sa matrice jacobienne $Jg(x,y,z)$ en (x,y,z).

2. Pour quels (x,y,z) la matrice $Jg(x,y,z)$ est-elle inversible ?

Exercice 3

Dans l'espace $\mathbb{R}^n$ muni de la norme $\|.\|$ on note $B(x,r) = \{y \in \mathbb{R}^n; \|y-x\| < r\}$ la boule ouverte de centre $x \in \mathbb{R}^n$ et rayon $r > 0$. Si A est un sous-ensemble de $\mathbb{R}^n$ on note

$$d(x,A) = \inf_{a \in A} \|x - a\|$$

la distance du point $x \in \mathbb{R}^n$ à A, et

$$\overline{A} = \{x \in \mathbb{R}^n; \forall r > 0, B(x,r) \cap A \neq \emptyset\}$$

l'adhérence de A.

1. a) Montrer que $\overline{A}$ est le plus petit fermé contenant A.

b) Montrer que $d(x,A) = 0$ si et seulement si $x \in \overline{A}$.

2. a) Pour tous x et y dans $\mathbb{R}^n$, montrer que

$$|d(x,A) - d(y,A)| \leq \|x - y\|.$$

b) En déduire que l'application f_A définie sur $\mathbb{R}^n$ par $f_A(x) = d(x,A)$ est continue sur $\mathbb{R}^n$.

c) Montrer que pour tout entier p strictement positif, l'ensemble

$$A_p = \left\{x \in \mathbb{R}^n; d(x,A) < \frac{1}{p}\right\}$$

est un ouvert de $\mathbb{R}^n$.

d) L'ensemble $\displaystyle\bigcap_{p=1}^{n} A_p$ est-il ouvert ou fermé ?

e) Montrer que $\displaystyle\bigcap_{p \in \mathbb{N}^*} A_p = \overline{A}$.

f) Donner un exemple de famille d'ensembles ouverts dont l'intersection est fermée.

3. Soient A et B deux fermés disjoints de $\mathbb{R}^n$.

a) Montrer que $d(x,A) + d(x,B) > 0$ pour tout x dans $\mathbb{R}^n$.

b) Montrer que la fonction g définie par

$$g(x) = \frac{f_A(x)}{f_A(x) + f_B(x)} = \frac{d(x,A)}{d(x,A) + d(x,B)}$$

est bien définie et continue sur $\mathbb{R}^n$. Vérifier qu'elle prend la valeur 0 sur A et 1 sur B.

c) En déduire l'existence de deux ouverts disjoints $\tilde{A}$ et $\tilde{B}$ de $\mathbb{R}^n$ contenant A et B respectivement.

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Contrôle continu du 21 mai 2007

Exercice 1

1) a) Soit f une fonction de classe C^2 sur un intervalle ouvert I de $\mathbb{R}$ à valeurs dans $\mathbb{R}$, concave et strictement positive.

Montrer que la fonction $g = \dfrac{1}{f}$ est convexe en calculant sa dérivée seconde $g''(x)$ pour $x \in I$.

b) Montrer que la fonction $f(x) = e^x$ est convexe sur $\mathbb{R}$. Qu'en est-il de son inverse $\dfrac{1}{f}$?

2) Soient f une fonction concave définie sur un sous-ensemble convexe U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$, et g une fonction concave croissante définie sur un intervalle I de $\mathbb{R}$ à valeurs dans $\mathbb{R}$, tels que $f(U) \subset I$. Montrer que $g \circ f$ est concave sur U.

3) Soit f une fonction concave à valeurs strictement positives définie sur un sous-ensemble convexe U de $\mathbb{R}^n$.

a) Montrer que la fonction $\ln(f)$ est concave sur U.

b) En déduire que la fonction $\dfrac{1}{f}$ est convexe sur U.

4) Soit $D = \{(x_1, \ldots, x_n) \in \mathbb{R}^n ; \forall i \in \{1, \ldots, n\}, x_i > 0.\}$.

a) Montrer que D est un sous-ensemble ouvert de $\mathbb{R}^n$.

b) Montrer que D est un sous-ensemble convexe de $\mathbb{R}^n$.

c) Montrer que la fonction f définie sur D par $f(x_1, \ldots, x_n) = \dfrac{1}{\displaystyle\sum_{i=1}^{n} \sqrt{x_i}}$ est convexe sur D.

5) Soient α et β deux réels non nuls, $E = \{(x, y) \in \mathbb{R}^2 ; x > 0, y > 0\}$, et f la fonction définie sur E par

$$f(x, y) = x^\alpha\, y^\beta.$$

On admet que E est un sous-ensemble ouvert et convexe de $\mathbb{R}^2$ et que la fonction f est de classe C^2 sur E.

a) Donner une condition nécessaire sur α et β pour que f soit convexe ou concave sur E.

b) Montrer que f est convexe sur E si $\alpha < 0$ et $\beta < 0$.

c) Montrer que f est concave sur E si $\alpha > 0, \beta > 0$ et $\alpha + \beta \leq 1$.

d) En déduire un exemple de fonction convexe strictement positive sur E dont l'inverse est concave sur E.

e) En déduire un exemple de fonction convexe strictement positive sur E dont l'inverse n'est ni convexe ni concave sur E.

Exercice 2

Montrer que la fonction f définie sur $\mathbb{R}^2$ par

$$f(x,y) = x^2 e^y - y e^x$$

est de classe C^2 sur $\mathbb{R}^2$, et écrire son développement limité d'ordre 2 au voisinage du point $(0,0)$.

Exercice 3

Soit f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = < Bx, x > + < c, x > + d, \qquad x \in \mathbb{R}^n$$

où B est une matrice symétrique de taille (n,n), c un vecteur de $\mathbb{R}^n$ et d un nombre réel.

Développer l'expression $f(x+h)$ pour $x, h \in \mathbb{R}^n$ et en déduire le développement limité à l'ordre 2 de f au voisinage de x.

Déterminer $\nabla f(x)$ et le reste du développement et montrer que le terme d'ordre 2 est indépendant de x.

Université Paris-Dauphine
DUMI2E 2e année
Calcul différentiel

Examen du 12 juin 2007

Le sujet comporte 2 pages. L'épreuve dure 2 heures. Les documents, calculatrices et téléphones portables sont interdits.

Exercice 1

Soit f la fonction définie sur $\mathbb{R}^2$ par $f(x,y) = \dfrac{-x^2-y^2}{|x|+|y|}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$.

1) Montrer que f est continue sur $\mathbb{R}^2$ et n'admet pas de dérivées partielles d'ordre 1 en $(0,0)$.

2) Montrer que le point $(0,0)$ est un point de maximum global strict de la fonction f.

Exercice 2

Soit φ la fonction définie sur $\mathbb{R}$ par $\varphi(x) = x\,e^{-x^2/2}$ pour $x \in \mathbb{R}$ et f la fonction définie sur $\mathbb{R}^2$ par $f(x,y) = xye^{-(x^2+y^2)/2}$ pour $(x,y) \in \mathbb{R}^2$.

1) Déterminer les extrema de φ sur $\mathbb{R}$ et préciser leur nature (maximum, minimum, local, global).

2) Déterminer les extrema de f sur $\mathbb{R}^2$ et préciser leur nature (maximum, minimum, local, global).

Dans les exercices 3 à 5 on munit l'espace $\mathbb{R}^n$ du produit scalaire euclidien $< \cdot, \cdot >$ et de la norme euclidienne $\| \cdot \|_2$.

Exercice 3

1) Montrer que la fonction F définie sur $\mathbb{R}^2$ par

$$F(x,y) = 2x^2 + y^2 - \sqrt{3}\,xy + 2x + 1\,, \qquad (x,y) \in \mathbb{R}^2$$

admet un unique point de minimum global sur $\mathbb{R}^2$ que l'on déterminera.

2) De façon générale, soit f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = < Ax, x > -2 < b, x > + c\,, \qquad x \in \mathbb{R}^n$$

où A est une matrice symétrique semi-définie positive de taille (n,n), b un vecteur de $\mathbb{R}^n$ et c un nombre réel.

a) Montrer que le gradient de f est donné par $\nabla f(x) = 2(Ax - b)$ et sa matrice hessienne par $Hf(x) = 2A$ pour $x \in \mathbb{R}^n$.

b) Montrer que f est une fonction convexe sur $\mathbb{R}^n$.

c) Montrer qu'un point x de $\mathbb{R}^n$ est un point de minimum global de f dans $\mathbb{R}^n$ si et seulement s'il est solution du système linéaire $Ax = b$.

d) Si de plus la matrice A est définie positive, montrer qu'il existe un unique point de minimum global de f dans $\mathbb{R}^n$ par chacune des deux méthodes suivantes :

i) en montrant que f est coercive,

ii) en utilisant le résultat de la question c).

Exercice 4

1) Montrer que la fonction F définie sur $\mathbb{R}^n$ par

$$F(x) = \sum_{i=1}^{n} x_i^2 - n\,, \qquad x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

admet un unique point de minimum global, que l'on déterminera, sur l'ensemble des contraintes

$$U = \{(x_1, \ldots, x_n) \in \mathbb{R}^n;\ \sum_{i=1}^{n} x_i = n\}.$$

2) De façon générale, soit f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = < Ax, x > -2 < b, x > + c\,, \qquad x \in \mathbb{R}^n$$

où A est une matrice symétrique définie positive de taille $(n, n), b$ un vecteur de $\mathbb{R}^n$ et c un nombre réel, et soit U l'ensemble des contraintes

$$U = \{x \in \mathbb{R}^n;\ < d, x >= e\}$$

où d est un vecteur non nul de $\mathbb{R}^n$ et e un nombre réel.

a) Montrer que f admet un unique point de minimum global dans U.

b) Montrer que U est un sous-ensemble convexe et fermé de $\mathbb{R}^n$.

Exercice 5

1) Montrer que la fonction F définie sur $\mathbb{R}^2$ par

$$F(x, y) = 2x^2 + y^2 - \sqrt{3}\,xy\,, \qquad (x, y) \in \mathbb{R}^2$$

admet au moins un point de maximum global sur le cercle unité $U = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$.

En déduire l'existence d'une valeur propre $\lambda \in \mathbb{R}$ et d'un vecteur propre associé $a \in \mathbb{R}^2$ de la matrice représentative A de la forme quadratique F; expliciter un tel couple (λ, a).

2) De façon générale, soit A une matrice symétrique de taille (n, n) et f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = < Ax, x >\,, \qquad x \in \mathbb{R}^n.$$

Montrer qu'il existe au moins un point a de maximum global de la fonction f sur la sphère unité $U = \{\, x \in \mathbb{R}^n,\ \|x\|_2 = 1\}$.

En déduire qu'il existe $\lambda \in \mathbb{R}$ tel que $Aa = \lambda a$.

Corrigé de l'examen du 12 juin 2007

Exercice 1

1) f est continue hors de $(0,0)$, et en $(0,0)$ on remarque que $|f(x,y)-f(0,0)| \leq |x|+|y|$ pour tout (x,y), donc f est continue en $(0,0)$.

L'application partielle par rapport à x, en $y=0$ fixé, est $x \mapsto -|x|$, qui n'est pas dérivable en 0 : autrement dit f n'admet pas de dérivée partielle par rapport à x en (0,0). De même pour y.

2) Pour tout $(x,y) \neq (0,0)$ on a $f(x,y) < 0 = f(0,0)$, donc le point $(0,0)$ est un point de maximum global strict de f.

Exercice 2

1) La fonction φ est dérivable sur $\mathbb{R}$ avec

$$\varphi'(x) = (1-x^2)\, e^{-x^2/2}.$$

Ainsi les éventuels points d'extremum de φ dans $\mathbb{R}$ sont parmi ses deux points critiques -1 et $+1$. Comme de plus

$$\varphi'(x) < 0 \quad \text{pour} \quad -\infty < x < -1$$
$$\varphi'(x) > 0 \quad \text{pour} \quad -1 < x < +1$$
$$\varphi'(x) < 0 \quad \text{pour} \quad +1 < x < +\infty$$

il en résulte que φ est strictement décroissante dans les intervalles $]-\infty,-1[$ et $]+1,\infty[$, et strictement croissante dans l'intervalle $]-1,+1[$.

Par conséquent les extrema de φ dans $\mathbb{R}$ sont d'une part le point -1 qui est l'unique point de minimum global de φ dans $\mathbb{R}$, avec $\varphi(-1) = -1/\sqrt{e}$ et d'autre part le point $+1$ qui est l'unique point de maximum global de φ dans $\mathbb{R}$, avec $\varphi(+1) = +1/\sqrt{e}$.

2) Comme $f(x,y) = \varphi(x)\,\varphi(y)$, la fonction f est différentiable sur $\mathbb{R}^2$ avec

$$\frac{\partial f}{\partial x}(x,y) = \varphi'(x)\,\varphi(y)$$
$$\frac{\partial f}{\partial y}(x,y) = \varphi(x)\,\varphi'(y).$$

Ainsi les éventuels extrema de f sont parmi ses cinq points critiques $(0,0)$ et (x,y) avec $|x| = |y| = 1$.

- Le point $(0,0)$ est un col car

$$f(0,0) = 0 < f(x,y) \quad \text{pour} \quad xy > 0$$
$$f(0,0) = 0 > f(x,y) \quad \text{pour} \quad xy < 0.$$

- Les points $(-1,-1)$ et $(1,1)$ sont des points de maximum global de f sur $\mathbb{R}^2$ car $f(-1,-1) = f(1,1) = 1/e$ et $f(x,y) = \varphi(x)\,\varphi(y) \in [-1/e,+1/e]$ pour tout $(x,y) \in \mathbb{R}^2$ puisque $\varphi(x) \in [-1/\sqrt{e},+1/\sqrt{e}]$ pour tout $x \in \mathbb{R}$.

- Les points $(-1,1)$ et $(1,-1)$ sont des points de minimum global de f sur $\mathbb{R}^2$ car $f(-1,1) = f(1,-1) = -1/e$. $\diamond$

Exercice 3

1) La fonction F est deux fois différentiable sur $\mathbb{R}^2$ avec

$$\nabla F(x,y) = (4x - \sqrt{3}y + 2, -\sqrt{3}x + 2y)$$

$$HF(x,y) = \begin{bmatrix} 4 & -\sqrt{3} \\ -\sqrt{3} & 2 \end{bmatrix}.$$

Si (x,y) est un point d'extremum de F, alors (x,y) est nécessairement solution de l'équation d'Euler, c'est-à-dire du système linéaire

$$\begin{cases} 4x - \sqrt{3}y = -2 & (1) \\ -\sqrt{3}x + 2y = 0 & (2) \end{cases}$$

Or de l'équation (2) on a

$$x = \frac{2}{\sqrt{3}}y$$

et de l'équation (1) on a

$$(\frac{8}{\sqrt{3}} - \sqrt{3})y = -2,$$

d'où

$$x = -\frac{4}{5} \qquad y = -\frac{2\sqrt{3}}{5}.$$

Ainsi le point $(-\frac{4}{5}, -\frac{2\sqrt{3}}{5})$ est l'unique point critique de F. Or la matrice hessienne de F est en tout point définie positive d'après la règle de Sylvester puisque ses deux déterminants mineurs sont égaux à 4 et 5. Donc F est strictement convexe dans $\mathbb{R}^2$. Par suite le point $(-\frac{4}{5}, -\frac{2\sqrt{3}}{5})$ est l'unique point de minimum global de la fonction F sur $\mathbb{R}^2$.

2) a) La fonction f est deux fois différentiable sur $\mathbb{R}^n$ avec $\nabla f(x) = 2(Ax - b)$ et $Hf(x) = 2A$ pour tout $x \in \mathbb{R}^n$.

b) Comme $Hf(x) = 2A$ pour tout $x \in \mathbb{R}^n$ et que la matrice A est semi-définie positive alors f est convexe dans $\mathbb{R}^n$.

c) Comme f est différentiable et convexe dans $\mathbb{R}^n$, un point x de $\mathbb{R}^n$ est un point de minimum local de f dans $\mathbb{R}^n$ si et seulement s'il est un point de minimum global de f dans $\mathbb{R}^n$ si et seulement s'il est un point critique de f dans $\mathbb{R}^n$, c'est-à-dire vérifie l'équation d'Euler $\nabla f(x) = 0$, soit $Ax - b = 0$.

d) Si la matrice A est définie positive alors f est strictement convexe dans $\mathbb{R}^n$, donc il existe au plus un point de minimum global de f sur $\mathbb{R}^n$. De plus

(i) comme la matrice A est définie positive, alors il existe $\alpha > 0$ telle que

$$< Ax, x > \geq \alpha \|x\|_2^2$$

d'où

$$f(x) \geq \alpha \|x\|_2^2 - 2\|b\|_2 \|x\|_2 - |c|.$$

Ainsi $f(x)$ tend vers $+\infty$ quand $\|x\|_2$ tend vers $+\infty$ et donc f est coercive dans $\mathbb{R}^n$; par suite il existe au moins un point de minimum global de f sur $\mathbb{R}^n$ par la théorème de Weierstrass.

(ii) comme la matrice A est définie positive, alors elle est inversible, donc le système linéaire $Ax = b$ admet une (unique) solution x dans $\mathbb{R}^n$ et donc il existe un (et un seul) point de minimum global de f dans $\mathbb{R}^n$ par la question c). $\diamond$

Exercice 4

1) Posant $g(x) = \sum_{i=1}^{n} x_i - n$, on a $\nabla g(x) = (1, \ldots, 1)$ et par conséquent chaque point x de $\mathbb{R}^n$ est un point de qualification de la contrainte g.

On cherche donc les éventuels points d'extremum de F sur U en cherchant les points critiques (x, λ) du lagrangien

$$L(x, \lambda) = F(x) - \lambda\, g(x)$$

dans $\mathbb{R}^n \times \mathbb{R}$ par la résolution de l'équation d'Euler

$$\nabla L(x, \lambda) = 0$$

soit

$$\begin{cases} 2x_i - \lambda = 0, \quad i = 1, \ldots, n \\ \displaystyle\sum_{i=1}^{n} x_i - n = 0. \end{cases}$$

Or ce système de $n+1$ équations à $n+1$ inconnues (x, λ) a une et une seule solution donnée par $x_i = 1$ pour $i = 1, \ldots, n$ et $\lambda = 2$.

Ainsi il existe un et un seul point critique du lagrangien dans $\mathbb{R}^n \times \mathbb{R}$, à savoir le point $((1, \ldots, 1), 2)$. Or la fonction F est convexe dans $\mathbb{R}^n$ et la contrainte g est affine, donc $L(., \lambda)$ est convexe, et donc le point

$$(1, \ldots, 1)$$

est un point de minimum global de F dans U. De plus par la condition nécessaire de Lagrange (puisque tous les points de U sont de qualification) il est le seul point d'extremum de F dans U.

2) a) Posant $g(x) =< d, x > -e$, on a $\nabla g(x) = d \neq 0$ et par conséquent chaque point x de $\mathbb{R}^n$ est un point de qualification de la contrainte g. On cherche donc les éventuels points d'extremum de f dans U en cherchant les points critiques (x, λ) du lagrangien $L(x, \lambda) = f(x) - \lambda\, g(x)$ dans $\mathbb{R}^n \times \mathbb{R}$ par la résolution des équations d'Euler de L dans $\mathbb{R}^n \times \mathbb{R}$

$$\nabla L(x, \lambda) = 0$$

soit

$$\begin{cases} 2Ax - \lambda\, d = 2b & (1) \\ < d, x >= e & (2). \end{cases}$$

Or ce système linéaire de $n+1$ équations à $n+1$ inconnues (x, λ) a une et une seule solution si et seulement si le système homogène associé

$$\begin{cases} 2Ax - \lambda\, d = 0 & (1') \\ < d, x >= 0 & (2') \end{cases}$$

a la seule solution $x = 0$, $\lambda = 0$. Or de (1') on a

$$2 < Ax, x > - < \lambda\, d, x >= 0$$

soit

$$2 < Ax, x > -\lambda < d, x >= 0$$

et avec (2') on a

$$< Ax, x >= 0.$$

Comme A est définie positive, on en déduit que $x = 0$, et revenant à (1') on a

$$\lambda\, d = 0.$$

Comme d est non nul, on en déduit que $\lambda = 0$.

Ainsi il existe un et un seul point critique $(a, \lambda) = (0, 0)$ du lagrangien L dans $\mathbb{R}^n \times \mathbb{R}$. Or la fonction f est différentiable et convexe dans $\mathbb{R}^n$ et la contrainte g est affine, donc $L(., \lambda)$ est convexe, et par conséquent le point a est un point de minimum global de f dans U. De plus par la condition nécessaire de Lagrange (puisque tous les points de U sont de qualification) il est le seul point d'extremum de f dans U. $\diamond$

2) b) Si x et y sont deux points de U, alors $< d, x >=< d, y >= e$, donc $< d, tx + (1 - t)y >= e$ pour tout $t \in [0, 1]$, c'est-à-dire $tx + (1-t)y \in U$. Autrement dit U est convexe.

Si $(x_n)_n$ est une suite de U convergeant vers un point x de $\mathbb{R}^n$, alors $(< d, x_n >)_n$ est une suite constante égale à e, et converge vers $< x, e >$ par continuité du produit scalaire : ainsi $< d, x >= e$, c'est-à-dire $x \in U$. Ainsi U est fermé.

Exercice 5

1) La fonction F est continue sur $\mathbb{R}^2$ et la sphère unité U est fermée et bornée dans $\mathbb{R}^2$, donc il existe au moins un point a de $\mathbb{R}^2$ qui est un point de maximum global de F sur U.

Posant $g(x, y) = x^2 + y^2 - 1$ on a $\nabla g(x, y) = 2(x, y)$ qui est non nul sur U, donc tous les points de U sont de qualification de la contrainte g. De plus la fonction F est différentiable dans $\mathbb{R}^2$. Par conséquent par la condition nécessaire de Lagrange, il existe λ dans $\mathbb{R}$ solution de l'équation de Lagrange associée

$$\nabla F(a) - \lambda\, \nabla g(a) = 0$$

Or si A est la matrice représentative de f dans la base canonique de $\mathbb{R}^2$, c'est-à-dire si $F(x, y) =< A(x, y), (x, y) >$ pour tout $x, y) \in \mathbb{R}^2$ on a $\nabla F(x, y) = 2A(x, y)$. Par conséquent l'équation de Lagrange précédente s'écrit

$$Aa = \lambda a,$$

autrement dit λ est une valeur propre de la matrice A associée au vecteur propre a normalisé par $\|a\|_2 = 1$.

Pour chercher explicitement les points (x, y) d'extremum de F dans U on cherche les points critiques $((x, y), \lambda)$ du Lagrangien L dans $\mathbb{R}^2 \times \mathbb{R}$, c'est-à-dire les points $((x, y), \lambda)$ solutions des équations d'Euler-Lagrange

$$\begin{cases} 4x - \sqrt{3}y - 2x\lambda = 0 & (1) \\ 2y - \sqrt{3}x - 2y\lambda = 0 & (2) \\ x^2 + y^2 - 1 = 0 & (3) \end{cases}$$

De l'équation (1) on a nécessairement

$$2(2-\lambda)x = \sqrt{3}y$$

et par suite par l'équation (2) on a

$$[-3 + 4(1-\lambda)(2-\lambda)]x = 0.$$

Ainsi
- soit $x = 0$ et par suite $y = 0$, ce qui est impossible par (3),
- soit $\lambda^2 - 3\lambda + \frac{5}{4} = 0$ et par suite $\lambda = \frac{5}{2}$ ou $\lambda = \frac{1}{2}$.

Pour $\lambda = \frac{5}{2}$, alors par (1) on a $x = -\sqrt{3}y$ et par suite par (3) on a $y = \pm\frac{1}{2}$. Ainsi à $\lambda = \frac{5}{2}$ correspondent les deux points $(\pm\frac{\sqrt{3}}{2}, \mp\frac{1}{2})$.

De même à $\lambda = \frac{1}{2}$ correspondent les deux points $(\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$.

Ainsi nécessairement si a est un point d'extremum de F dans U, alors

$$a = (\pm\frac{\sqrt{3}}{2}, \mp\frac{1}{2}) \qquad \text{ou} \qquad a = (\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2}).$$

Or

$$F(\pm\frac{\sqrt{3}}{2}, \mp\frac{1}{2}) = \frac{5}{2} \qquad \text{et} \qquad F(\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2}) = \frac{1}{2}$$

et comme il existe au moins un point de maximum global de F dans U et un point de minimum global de F dans U, on en déduit que
- les points $(\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$ sont des points de minimum global de F dans U égal à $\frac{5}{2}$,
- les points $(\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$ sont des points de minimum global de F dans U égal à $\frac{1}{2}$.

En conclusion par exemple $\lambda = \frac{5}{2}$ est une valeur propre de A associé au vecteur propre $(\frac{\sqrt{3}}{2}, -\frac{1}{2})$.

2) On cherche à maximiser la fonction quadratique f définie sur $\mathbb{R}^n$ par $f(x) =< Ax, x >$ sur la sphère unité $U = \{ x \in \mathbb{R}^n, \ \|x\|_2 = 1\}$.

La fonction f est continue de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et U est fermé et borné dans $\mathbb{R}^n$. Donc par le théorème de Weierstrass, il existe au moins un point a de maximum global de f sur U.

Posant $g(x) = \|x\|_2^2$, cette fonction g est différentiable dans $\mathbb{R}^n$ avec $\nabla g(x) = 2x$ qui est non nul dans U, donc tous les points de U sont des points de qualification de la contrainte. De plus la fonction f est différentiable sur $\mathbb{R}^n$ avec $\nabla f(x) = 2A$, donc par la condition nécessaire de Lagrange, il existe un multiplicateur λ dans $\mathbb{R}$ tel que

$$2Aa = 2\lambda a,$$

soit

$$Aa = \lambda a.$$

Donc λ est une valeur propre de A associée au vecteur propre a de $\mathbb{R}^n$ normalisé par $\|a\|_2 = 1$.

Université Paris-Dauphine
DUMI2E 2e année
Calcul différentiel

Examen de septembre 2007 - Deuxième session

Le sujet comporte 2 pages. L'épreuve dure 2 heures. Les documents, calculatrices et téléphones portables sont interdits.

Exercice 1
Soit f une fonction continue de $\mathbb{R}^n$ dans $\mathbb{R}^p$ et O un sous-ensemble ouvert de $\mathbb{R}^p$. Montrer que $f^{-1}(O)$ est un sous-ensemble ouvert de $\mathbb{R}^n$.

Exercice 2
Soit K un sous-ensemble fermé et borné de $\mathbb{R}^n$ et f une fonction continue sur K à valeurs réelles strictement positives. Montrer qu'il existe un nombre réel strictement positif m tel que $f(x) \geq m$ pour tout $x \in K$.
Le résultat reste-il vrai si K est fermé mais pas borné ? si K est borné mais pas fermé ?

Exercice 3
Soit f la fonction définie sur $\mathbb{R}^2$ par

$$f(x,y) = xy, \qquad (x,y) \in \mathbb{R}^2.$$

1) La fonction f admet-elle des extrema sur $\mathbb{R}^2$?
2) Déterminer les extrema de f sur l'ensemble des contraintes $U = \{(x,y) \in \mathbb{R}^2; x^4+y^4 = 2\}$ et préciser leur nature (maximum, minimum, local, global).

Exercice 4
Soit n points (x_i, y_i) de $\mathbb{R}^2$, pour $i = 1, \ldots, n$ et f la fonction définie sur $\mathbb{R}^2$ par

$$f(x,y) = \sum_{i=1}^{n} (x-x_i)^2 + (y-y_i)^2, \qquad (x,y) \in \mathbb{R}^2.$$

1) Déterminer les extrema de f sur $\mathbb{R}^2$ et préciser leur nature (maximum, minimum, local, global).
2) Déterminer les extrema de f sur l'ensemble des contraintes $U = \{(x,y) \in \mathbb{R}^2; x+y = 1\}$ et préciser leur nature (maximum, minimum, local, global).

Dans les exercices 5 à 7 on munit l'espace $\mathbb{R}^n$ du produit scalaire euclidien $< \cdot, \cdot >$ et de la norme euclidienne $\| \cdot \|_2$. Si M est une matrice de taille (n,n) on note M^T sa matrice transposée, c'est-à-dire la matrice de taille (n,n) définie par $< M^T y, z > = < y, Mz >$ pour $y, z \in \mathbb{R}^n$.

Exercice 5

Soit $D = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n; x_i > 0, i = 1, \ldots, n\}$ et f la fonction définie sur D par

$$f(x) = x_1^{1/n} \ldots x_n^{1/n}, \qquad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

1) Montrer que D est un sous-ensemble ouvert et convexe de $\mathbb{R}^n$.

2) Montrer que la fonction f est de classe C^2 sur D puis calculer son gradient ∇f et sa matrice hessienne Hf.

3) Montrer que

$$< Hf(x)h, h >= \frac{f(x)}{n^2}\Big[\Big(\sum_{i=1}^{n} \frac{h_i}{x_i}\Big)^2 - n\sum_{i=1}^{n}\Big(\frac{h_i}{x_i}\Big)^2\Big]$$

pour tous $x \in D$ et $h \in \mathbb{R}^n$.

4) En déduire que f est concave sur D (on pourra utiliser l'inégalité de Cauchy-Schwarz).

5) la fonction f admet-elle des extrema sur D ?

Exercice 6

Soit a et b deux points donnés de $\mathbb{R}^n$ avec $a + b \neq 0$ et f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = \|x - a\|_2^2 + \|x - b\|_2^2, \qquad x \in \mathbb{R}^n$$

1) Montrer que la fonction f est strictement convexe sur $\mathbb{R}^n$.

2) Montrer que la fonction f est coercive sur $\mathbb{R}^n$.

3) Montrer que la fonction f admet un unique minimum gobal sur $\mathbb{R}^n$ que l'on déterminera.

4) Déterminer les extrema de la fonction f sur la sphère de $\mathbb{R}^n$ de centre 0 et rayon 1 et préciser leur nature. Que se passe-t-il si $a + b = 0$?

Exercice 7

Soit f la fonction définie sur $\mathbb{R}^n$ par

$$f(x) = \|Bx - d\|_2^2, \qquad x \in \mathbb{R}^n$$

où B est une matrice de taille (n, n) et d un vecteur de $\mathbb{R}^n$.

1) Montrer que le gradient de f est donné par $\nabla f(x) = 2(B^T Bx - B^T d)$ et sa matrice hessienne par $Hf(x) = 2B^T B$ pour $x \in \mathbb{R}^n$.

2) Montrer que la fonction f est convexe sur $\mathbb{R}^n$.

3) Montrer qu'un point x de $\mathbb{R}^n$ est un point de minimum global de f sur $\mathbb{R}^n$ si et seulement s'il est solution du système linéaire $B^T Bx = B^T d$, par chacune des trois méthodes suivantes :

a) en appliquant la question 2),

b) en développant l'expression $f(x + th)$ sous forme d'un polynôme d'ordre 2 en la variable réelle t,

c) en appliquant au point d de $\mathbb{R}^n$ le théorème de projection sur l'image de B.

4) Si de plus B est inversible, montrer que f admet un unique point de minimum global sur $\mathbb{R}^n$.

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Contrôle continu du 25 mars 2008

Le sujet comporte 2 pages. L'épreuve dure 1 heure 30. Les documents, calculatrices et téléphones portables sont interdits.

Questions de cours

1. Donner la définition d'un ensemble ouvert de $\mathbb{R}^n$.
2. Montrer que l'union d'une famille quelconque d'ouverts de $\mathbb{R}^n$ est un ouvert de $\mathbb{R}^n$.
3. Montrer que l'intersection de deux ouverts de $\mathbb{R}^n$ est un ouvert de $\mathbb{R}^n$.
4. L'intersection d'une famille quelconque d'ouverts de $\mathbb{R}^n$ est-il un ouvert de $\mathbb{R}^n$?

Exercice 1

Soit f la fonction définie sur $]-1,1[\times]-1,1[$ par

$$f(x,y) = \begin{cases} \dfrac{x^2-y^2}{x^2+y^2} + \ln(1-x^2) + \ln(1-y^2) & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

1. Montrer que $]-1,1[\times]-1,1[$ est un ouvert de $\mathbb{R}^2$.
2. Montrer que f est continue sur $]-1,1[\times]-1,1[\setminus\{(0,0)\}$.
3. Calculer $\lim\limits_{y\to 0, y\neq 0} f(x,y)$ pour $x \neq 0$ fixé, puis $\lim\limits_{x\to 0, x\neq 0}\big[\lim\limits_{y\to 0, y\neq 0} f(x,y)\big]$.
4. Calculer $\lim\limits_{x\to 0, x\neq 0} f(x,y)$ pour $y \neq 0$ fixé, puis $\lim\limits_{y\to 0, y\neq 0}\big[\lim\limits_{x\to 0, x\neq 0} f(x,y)\big]$.
5. Etudier la continuité de la restriction de f à chaque droite passant par 0.
6. Etudier la continuité de la fonction f en $(0,0)$.

Exercice 2

Soit $B(0,r)$ la boule ouverte de $\mathbb{R}^n$ de centre 0 et de rayon $r > 0$ pour une norme $\|\cdot\|$ sur $\mathbb{R}^n$.

Soit L un nombre réel de $]0,1[$ et f une application de $B(0,r)$ dans $\mathbb{R}^n$ telle que

$$\|f(x) - f(y)\| \leq L\,\|x - y\|$$

pour tous x et y dans $B(0,r)$ et telle que

$$\|f(0)\| \leq r(1 - L).$$

On va montrer qu'il existe un unique point x de $B(0,r)$ tel que $f(x) = x$. Un tel point est appelé un point fixe de l'application f.

1. Montrer que l'application f est continue sur $B(0,r)$.

2. Montrer qu'il existe au plus un point fixe (en supposant que x et y sont deux points fixes, on pourra montrer que $x = y$.)

3. a. On note $x_0 = 0$ et $x_1 = f(x_0)$. Montrer que

$$\|x_1 - x_0\| \leq r(1 - L).$$

b. De façon générale montrer que l'on peut définir de proche en proche la suite $(x_k)_k$ de points de $B(0,r)$ par $x_{k+1} = f(x_k)$ pour $k \in \mathbb{N}$ et que pour tout $k \in \mathbb{N}$ on a

$$\|x_{k+1} - x_k\| \leq L^k\,\|x_1 - x_0\|.$$

c. En déduire que

$$\|x_{k+p} - x_k\| \leq \frac{L^k}{1 - L}\,\|x_1 - x_0\|$$

pour tous k, p dans $\mathbb{N}$, puis que la suite $(x_k)_k$ est de Cauchy dans $\mathbb{R}^n$.

d. En déduire l'existence d'un point fixe de l'application f.

Exercice 3

Soit $B(0,r)$ la boule ouverte de $\mathbb{R}^n$ de centre 0 et de rayon r pour une norme $\|\cdot\|$ sur $\mathbb{R}^n$.

Soit L un nombre réel de $]0,1[$ et g une fonction de $B(0,r)$ dans $\mathbb{R}^n$ telle que

$$\|g(x) - g(y)\| \leq L\,\|x - y\|$$

pour tous x et y dans $B(0,r)$ et telle que

$$g(0) = 0.$$

1. Soit $z \in B(0, r(1-L))$ fixé. A l'aide de l'exercice 2, montrer qu'il existe $x \in B(0,r)$ tel que $z - g(x) = x$.

2. Pour x_1 et $x_2 \in B(0,r)$ montrer que

$$\|x_1 - x_2\| \leq \frac{1}{1 - L}\|(x_1 + g(x_1)) - (x_2 + g(x_2))\|.$$

3. Soit h la fonction de $B(0,r)$ dans $\mathbb{R}^n$ définie par $h(x) = x + g(x)$.

Déduire de ce qui précède que la fonction h est une bijection de $B(0,r)$ sur $B(0, r(1-L))$, qui est continue dans $B(0,r)$ et que son application inverse est continue dans $B(0, r(1-L))$.

Corrigé du contrôle continu du 25 mars 2008

Questions de cours

1. cours

2. Soit $(O_i)_{i\in I}$ une famille d'ouverts et soit $x \in \bigcup_{i\in I} O_i$ quelconque. Alors il existe $i \in I$ tel que $x \in O_i$. Comme cet O_i est ouvert, il existe un $r > 0$ et une norme sur $\mathbb{R}^n$ tels que la boule de centre x et rayon r pour cette norme soit contenue dans O_i. En particulier elle est contenue dans $\bigcup_{i\in I} O_i$, ce qui assure que $\bigcup_{i\in I} O_i$ est ouvert.

3. Soit U et V deux ouverts de $\mathbb{R}^n$, et soit $x \in U \cap V$ quelconque. Alors $x \in U$ qui est ouvert, donc il existe $r > 0$ et une norme sur $\mathbb{R}^n$ tels que la boule $B(x, r)$ de centre x et rayon r pour cette norme soit contenue dans U. De même $x \in V$ qui est ouvert, donc il existe $s > 0$ tel que la boule $B(x, s)$ pour la même norme soit contenue dans V. En particulier $B(x, \min(r, s))$ est contenue dans U et dans V, donc dans $U \cap V$, ce qui assure que $U \cap V$ est ouvert.

4. L'intersection d'une famille quelconque d'ouverts de $\mathbb{R}^n$ n'est pas nécessairement un ouvert de $\mathbb{R}^n$. Par exemple les ensembles $]-1/n, 1/n[$ sont des ouverts de $\mathbb{R}$, mais leur intersection est le singleton $\{0\}$ qui n'est pas ouvert.

Exercice 1

1. Le produit tensoriel $]-1,1[\times]-1,1[$ est un ouvert de $\mathbb{R}^2$ puisque par exemple il est la boule ouverte de centre $(0,0)$ et de rayon 1 pour la norme $\|\cdot\|_\infty$.

2. La fonction $x \mapsto 1 - x^2$ ne prend que des valeurs strictement positives sur $]-1,1[$ donc par composition $x \mapsto \ln(1-x^2)$ est continue sur $]-1,1[$. Par conséquent $(x,y) \mapsto \ln(1-x^2)$ est continue sur $]-1,1[\times]-1,1[$, et en particulier sur $]-1,1[\times]-1,1[\setminus\{(0,0)\}$.
De plus $(x,y) \mapsto \dfrac{x^2-y^2}{x^2+y^2}$ est le quotient de deux fonctions continues, le dénominateur n'étant jamais nul sur $\mathbb{R}^2 \setminus \{(0,0)\}$, donc est continue sur $\mathbb{R}^2 \setminus \{(0,0)\}$, et en particulier sur $]-1,1[\times]-1,1[\setminus\{(0,0)\}$. Par conséquent f est continue sur $]-1,1[\times]-1,1[\setminus\{(0,0)\}$.

3. Pour $x \neq 0$ fixé, $\lim\limits_{y\to 0, y\neq 0} f(x,y) = 1 + \ln(1-x^2)$ donc $\lim\limits_{x\to 0, x\neq 0} \big[\lim\limits_{y\to 0, y\neq 0} f(x,y)\big] = 1$.

4. De même $\lim\limits_{x\to 0, x\neq 0} f(x,y) = -1+\ln(1-y^2)$ pour $y \neq 0$ fixé, donc $\lim\limits_{y\to 0, y\neq 0} \big[\lim\limits_{x\to 0, x\neq 0} f(x,y)\big] = -1$.

5. f est continue sur $]-1,1[\times]-1,1[\setminus\{(0,0)\}$ donc sa restriction à toute droite passant par 0 est continue en dehors de 0.

De plus, sur la droite $y = ax$ pour $a \in \mathbb{R}$ fixé, on a $f(x,y) = f(x,ax) = \frac{1-a^2}{1+a^2} + \ln(1-x^2) + \ln(1-a^2x^2)$ pour $x \neq 0$, qui tend vers $\dfrac{1-a^2}{1+a^2}$ quand x tend vers 0. Comme $f(0,0) = 0$, on en déduit que la restriction de f à la droite $y = ax$ est continue en 0 ssi $a^2 = 1$.

Enfin, sur la droite $x = 0$, on a $f(x,y) = f(0,y) = -1 + \ln(1-y^2)$ qui tend vers -1 quand y tend vers 0. Comme $f(0,0) = 0$, on en déduit que la restriction de f à la droite $x = 0$ n'est pas continue en 0.

6. D'après la question 5. la restriction de f a au moins une droite passant par 0 n'est pas continue en 0 donc f n'est pas continue en 0.

Exercice 2

1. La fonction est lipschitzienne de rapport L dans $B(0,r)$ donc y est continue.

2. Si x et y sont deux points de $B(0,r)$ tels que $f(x)=x$ et $f(y)=y$, alors par hypothèse

$$\|x-y\| = \|f(x)-f(y)\| \leq L\|x-y\|$$

et comme $L<1$ on a nécessairement $\|x-y\|=0$, soit $x=y$. Ainsi il existe au plus un point fixe de f.

3. a. On a par hypothèse

$$\|x_1-x_0\| = \|f(0)\| \leq r(1-L).$$

b. D'après la question précédente on a $\|x_1-x_0\| = \|f(0)\| \leq r(1-L) < r$ donc $x_1 \in B(0,r)$. On peut donc définir $x_2 = f(x_1)$ De plus

$$\|x_1-x_0\| \leq L^0\|x_1-x_0\|$$
$$\|x_2-x_1\| = \|f(x_1)-f(x_0)\| \leq L^1\|x_1-x_0\|.$$

Soit maintenant k un entier ≥ 1 et supposons que pour tout $n \leq k$ on a $x_n \in B(0,r)$, ce qui permet de définir $x_{n+1} = f(x_n)$, et que $\|x_{n+1}-x_n\| \leq L^n\|x_1-x_0\|$.

Par l'inégalité triangulaire on a

$$\|x_{k+1}-x_0\| \leq \|x_{k+1}-x_k\| + \cdots + \|x_1-x_0\|$$

et par l'hypothèse de récurrence sur la majoration on en déduit que

$$\|x_{k+1}-x_0\| \leq (L^k+\cdots+L^0)\|x_1-x_0\| = \frac{1-L^{k+1}}{1-L}\,r(1-L) < r.$$

Ainsi $x_{k+1} \in B(0,r)$, donc on peut définir $x_{k+2} = f(x_{k+1})$, et de plus $\|x_{k+2}-x_{k+1}\| = \|f(x_{k+1})-f(x_k)\| \leq L\|x_{k+1}-x_k\| \leq L^{k+1}\|x_1-x_0\|$.

Ainsi la propriété est récurrente et par conséquent on peut bien définir de proche en proche la suite $(x_k)_k$ par $x_{k+1} = f(x_k)$ pour $k \geq 0$ avec la majoration annoncée.

c. Pour k et p dans $\mathbb{N}$ avec $p \geq 1$, on a par l'inégalité triangulaire puis les inégalités de la question précédente

$$\begin{aligned}\|x_{k+p}-x_k\| &\leq \|x_{k+p}-x_{k+p-1}\| + \cdots + \|x_{k+1}-x_k\| \leq \left(L^{k+p-1}+\cdots+L^k\right)\|x_1-x_0\| \\ &= L^k\left(L^{p-1}+\cdots+L^0\right)\|x_1-x_0\| \leq \frac{L^k}{1-L}\,\|x_1-x_0\|.\end{aligned}$$

On peut alors en déduire que la suite $(x_k)_k$ est de Cauchy dans $\mathbb{R}^n$. En effet on déduit tout d'abord de la majoration précédente et de a. que

$$\|x_{k+p}-x_k\| \leq L^k r.$$

Soit alors $\varepsilon > 0$ fixé. Comme la suite $(L^k)_k$ est convergente vers 0 dans $\mathbb{R}$, il existe un entier K tel que pour tout $k \geq K$ on ait

$$rL^k \leq \varepsilon.$$

Par conséquent pour tout entier $k \geq K$ et pour tout entier $p \geq 0$ on a

$$\|x_{k+p} - x_k\| \leq \varepsilon$$

et donc la suite $(x_k)_k$ est de Cauchy dans $\mathbb{R}^n$.

d. Il existe donc un point x de $\mathbb{R}^n$ limite de cette suite dans $\mathbb{R}^2$ et qui vérifie (par l'inégalité de la question précédente écrite pour $k = 0$ et p tendant vers l'infini)

$$\|x - x_0\| \leq \frac{1}{1-L} \|x_1 - x_0\|$$

soit

$$\|x - x_0\| \leq \frac{1}{1-L} r(1-L) < r$$

donc $x \in B(0, r)$.

Comme la suite $(x_k)_k$ converge vers x dans $B(0, r)$ et comme la fonction f est continue dans $B(0, r)$ donc en x, la suite $(f(x_k))_k$ converge vers $f(x)$ dans $\mathbb{R}^n$. Enfin comme $x_{k+1} = f(x_k)$, la suite $(f(x_k))_k$ est une suite extraite de la suite $(x_k)_k$ donc converge aussi vers x. Par unicité de la limite d'une suite, on en déduit que $x = f(x)$. Ainsi ce point x est un point fixe de f.

Exercice 3

1. La fonction f de $B(0, r)$ dans $\mathbb{R}^2$ définie par

$$f(x) = z - g(x)$$

vérifie

$$\|f(x) - f(y)\| = \|g(x) - g(y)\| \leq L\|x - y\|$$

pour tous x et $y \in B(0, r)$, et

$$\|f(0)\| = \|z - g(0)\| = \|z\| \leq r(1 - L).$$

Par l'exercice 2, il existe donc $x \in B(0, r)$ tel que $f(x) = x$, c'est-à-dire $z - g(x) = x$.

2. Des égalités

$$x_1 = x_1 + g(x_1) - g(x_1)$$
$$x_2 = x_2 + g(x_2) - g(x_2)$$

on déduit par différence

$$x_1 - x_2 = ((x_1 + g(x_1)) - (x_2 + g(x_2))) - (g(x_1) - g(x_2))$$

puis par inégalité triangulaire et hypothèse sur g

$$\|x_1 - x_2\| \leq \|(x_1 + g(x_1)) - (x_2 + g(x_2))\| + \|g(x_1) - g(x_2)\| \leq \|(x_1 + g(x_1)) - (x_2 + g(x_2))\| + L\|x_1 - x_2\|$$

d'où

$$\|x_1 - x_2\| \leq \frac{1}{1-L} \|(x_1 + g(x_1)) - (x_2 + g(x_2))\|.$$

3. L'hypothèse sur g assure que la fonction g puis la fonction h est continue sur $B(0, r)$ et la question 1 assure qu'elle est surjective de $B(0, r)$ sur $B(0, r(1 - L))$. La question 2 assure qu'elle est injective dans $B(0, r)$ avec une application inverse qui est continue.

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Examen partiel du 14 avril 2008

Questions de cours

1. Soit U un sous-ensemble fermé et borné de $\mathbb{R}^n$ et f une fonction définie et continue sur U à valeurs dans $\mathbb{R}^p$. Montrer que l'ensemble $f(U)$ est un sous-ensemble fermé et borné de $\mathbb{R}^p$.

2. Soit f une forme quadratique sur $\mathbb{R}^n$, représentée dans la base canonique de $\mathbb{R}^n$ par une matrice A symétrique de taille (n, n). Montrer que f est deux fois différentiable sur $\mathbb{R}^n$ et déterminer son gradient et sa matrice hessienne en fonction de la matrice A en développant l'expression $f(x+h)$ pour x et h dans $\mathbb{R}^n$.

Exercice 1

Soit f la fonction définie sur $\mathbb{R}^2$ par

$$f(x,y) = \begin{cases} \dfrac{x^3 y}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

1. Montrer que f est continue sur $\mathbb{R}^2$.

2. Montrer que f est une fois continûment différentiable (ou de classe C^1) sur $\mathbb{R}^2$ et calculer ses dérivées partielles premières.

3. Montrer que f admet des dérivées partielles secondes en $(0, 0)$ que l'on calculera.

4. L'application f est-elle deux fois continûment différentiable sur $\mathbb{R}^2$?

Exercice 2

Si A et B sont deux sous-ensembles non vides de $\mathbb{R}^n$, on note $A+B$ le sous-ensemble de $\mathbb{R}^n$ défini par

$$A + B = \{z \in \mathbb{R}^n ; z = a + b,\ a \in A,\ b \in B\}.$$

1. (Exemple dans $\mathbb{R}^2$) Si $A = \{(x, y) \in \mathbb{R}^2 ; y = 0\}$ et $B = \{(x, y) \in \mathbb{R}^2 ; x = 0\}$, déterminer $A + B$.

2. Montrer que si A et B sont bornés, alors $A + B$ est borné.

3. Montrer que si A et B sont convexes, alors $A + B$ est convexe.

4. On suppose que A est ouvert.

a. Montrer que $A + B = \bigcup_{b \in B} A + \{b\}$.

b. Montrer que $A + \{b\}$ est ouvert pour tout $b \in B$ et en déduire que $A + B$ est ouvert.

5. Montrer que si A est fermé borné et si B est fermé, alors $A + B$ est fermé (on pourra raisonner à l'aide de suites et suites extraites).

6. Montrer que si A et B sont fermés, alors $A+B$ n'est pas nécessairement fermé (on pourra considérer les sous-ensembles $A=\{(x,y)\in\mathbb{R}^2;\, y=0\}$ et $B=\{(x,y)\in\mathbb{R}^2;\, xy=1\}$ de $\mathbb{R}^2$).

Exercice 3

Soit f l'application définie sur $\mathbb{R}^2$ par $f(x,y)=e^{x^2+y^2}\Big(x^2+xy+y^2-\frac{1}{2}\Big)$.

1. Montrer que $a^2+ab+b^2\geq\frac{1}{2}(a^2+b^2)$ pour tous réels a et b.

2. Montrer que l'application f est minorée sur $\mathbb{R}^2$ et atteint sa borne inférieure. L'application f est-elle majorée sur $\mathbb{R}^2$?

3. Montrer que f est une fois continûment différentiable sur $\mathbb{R}^2$ et calculer ses dérivées partielles premières.

4. Soit (x_0,y_0) un point de minimum de f sur $\mathbb{R}^2$, c'est-à-dire tel que $f(x,y)\geq f(x_0,y_0)$ pour tout (x,y) de $\mathbb{R}^2$.

a. Montrer que l'application partielle $x\mapsto f(x,y_0)$ admet un minimum en x_0 et en déduire que $\dfrac{\partial f}{\partial x}(x_0,y_0)=0$.

b. Montrer de même que $\dfrac{\partial f}{\partial y}(x_0,y_0)=0$.

5. a. Déterminer tous les couples (x,y) de $\mathbb{R}^2$ solution du système

$$\begin{cases}\dfrac{\partial f}{\partial x}(x,y) & =0\\ \dfrac{\partial f}{\partial y}(x,y) & =0\end{cases}$$

(on pourra former la différence des deux équations).

b. En déduire qu'il existe un seul point de minimum de f sur $\mathbb{R}^2$ que l'on explicitera, ainsi que la valeur du minimum de f sur $\mathbb{R}^2$.

6. Si k est un nombre réel on appelle courbe de niveau k l'ensemble (éventuellement vide) $C(k)$ défini par

$$C(k)=\{(x,y)\in\mathbb{R}^2; f(x,y)=k\}.$$

a. Déterminer l'ensemble $C(-1/2)$.

b. Pour quelles valeurs de k l'ensemble $C(k)$ est-il vide ?

c. Montrer que l'ensemble $C(k)$ est un ensemble fermé et borné de $\mathbb{R}^2$ pour tout réel k.

Corrigé succinct de l'examen partiel du 14 avril 2008

Exercice 1

1. La fonction f est continue sur $\mathbb{R}^2 \setminus \{(0,0)\}$ comme quotient de fonctions continues avec un dénominateur non nul sur $\mathbb{R}^2 \setminus \{(0,0)\}$. De plus, pour $(x,y) \neq (0,0)$,

$$|f(x,y) - f(0,0)| = \frac{|x^3y|}{x^2+y^2} = \frac{|x^2|}{x^2+y^2}|xy| \leq |xy|$$

qui tend vers 0 quand (x,y) tend vers $(0,0)$, donc f est continue en $(0,0)$.

2. La fonction f admet des dérivées partielles sur l'ouvert $\mathbb{R}^2 \setminus \{(0,0)\}$ avec

$$\frac{\partial f}{\partial x}(x,y) = \frac{x^4y+3x^2y^3}{(x^2+y^2)^2} \qquad , \qquad \frac{\partial f}{\partial y}(x,y) = \frac{x^5-x^3y^2}{(x^2+y^2)^2}.$$

qui sont continues comme quotient de fonctions continues avec un dénominateur non nul sur $\mathbb{R}^2 \setminus \{(0,0)\}$.

L'application partielle $f(.,0) : x \mapsto f(x,0) = 0$ est dérivable en $x=0$, de dérivée 0, donc l'application f admet une dérivée partielle par rapport à x en $(0,0)$, avec $\frac{\partial f}{\partial x}(0,0) = 0$. De plus, pour $(x,y) \neq (0,0)$,

$$\left|\frac{\partial f}{\partial x}(x,y) - \frac{\partial f}{\partial x}(0,0)\right| = \frac{|x^4y+3x^2y^3|}{(x^2+y^2)^2} \leq |y|\frac{x^4}{(x^2+y^2)^2} + \frac{3}{2}|y|\frac{2x^2y^2}{(x^2+y^2)^2} \leq |y|(1+\frac{3}{2}).$$

La fonction $\frac{\partial f}{\partial x}$ est donc continue en $(0,0)$ et par suite continue sur $\mathbb{R}^2$.

On montre de même que f admet une dérivée partielle par rapport à y, continue sur $\mathbb{R}^2$.

Ainsi la fonction f est ainsi de classe C^1 sur $\mathbb{R}^2$ avec

$$\frac{\partial f}{\partial x}(x,y) = \frac{x^4y+3x^2y^3}{(x^2+y^2)^2} \quad \text{si} \quad (x,y) \neq (0,0), \quad 0 \quad \text{sinon}$$

$$\frac{\partial f}{\partial y}(x,y) = \frac{x^5-x^3y^2}{(x^2+y^2)^2} \quad \text{si} \quad (x,y) \neq (0,0), \quad 0 \quad \text{sinon.}$$

3. L'application partielle $\frac{\partial f}{\partial x}(0,.) : y \mapsto \frac{\partial f}{\partial x}(0,y) = 0$ est dérivable en $y=0$, de dérivée 0, donc $\frac{\partial^2 f}{\partial y \partial x}(0,0)$ existe et vaut 0. De même $\frac{\partial^2 f}{\partial x^2}(0,0) = 0$, $\frac{\partial^2 f}{\partial y^2}(0,0) = 0$ et $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$.

4. $\frac{\partial^2 f}{\partial y \partial x}(0,0) \neq \frac{\partial^2 f}{\partial x \partial y}(0,0)$ donc f n'est pas deux fois continûment différentiable sur $\mathbb{R}^2$ par le théorème de Schwarz.

Exercice 2

1. $A + B = \mathbb{R}^2$.

2. Soit $\|.\|$ une norme sur $\mathbb{R}^n$. Si A est borné, alors il existe une constante c_1 telle que $\|a\| \leq c_1$ pour tout $a \in A$. Si de même B est borné, alors il existe une constante c_2 telle que $\|b\| \leq c_2$ pour tout $b \in B$.

Ainsi, si $z \in A + B$, alors il existe $a \in A$ et $b \in B$ tels que $z = a + b$, et par inégalité triangulaire $\|z\| \leq \|a\| + \|b\| \leq c_1 + c_2$. Comme ceci est vrai pour tout $z \in A + B$ on en déduit que $A + B$ est borné.

3. Soit z_1 et z_2 deux points de $A + B$ et $t \in [0, 1]$: il existe donc des points $a_1, a_2 \in A$ et $b_1, b_2 \in B$ tels que $z_1 = a_1 + b_1$ et $z_2 = a_2 + b_2$. On considère alors

$$tz_1 + (1 - t)z_2 = t(a_1 + b_1) + (1 - t)(a_2 + b_2) = ta_1 + (1 - t)a_2 + tb_1 + (1 - t)b_2.$$

Or a_1 et a_2 sont deux points du convexe A, donc $ta_1 + (1-t)a_2 \in A$. De même $tb_1 + (1-t)b_2 \in B$, et donc $tz_1 + (1 - t)z_2 \in A + B$, ce qui assure que $A + B$ est convexe.

4. a. $A + B = \bigcup_{b \in B}\{a + b, a \in A\} = \bigcup_{b \in B} A + \{b\}$.

b. A est ouvert donc $A + \{b\}$ est ouvert pour tout $b \in B$. En particulier, et d'après la question a., $A + B$ est une réunion d'ensembles ouverts, donc est ouvert.

5. Soit $(z_k)_k$ une suite de $A + B$, convergeant vers un $z \in \mathbb{R}^n$. Pour tout k, $z_k \in A + B$ donc il existe $a_k \in A$ et $b_k \in B$ tels que $z_k = a_k + b_k$. La suite $(a_k)_k$ ainsi définie appartient au fermé borné A, donc admet une sous-suite $(a_{k'})_{k'}$ convergeant vers un $a \in A$. La suite $(b_{k'})_{k'} = (z_{k'} - a_{k'})_{k'}$ converge donc vers $z - a$. Or elle est composée d'éléments du fermé B, donc la limite $z - a$ appartient à B. La limite z de la suite $(z_k)_k$ s'écrit donc comme $a + z - a$ où $a \in A$ et $z - a \in B$: autrement dit $z \in A + B$, ce qui assure que $A + B$ est fermé.

6. L'ensemble $A = \{(x, y) \in \mathbb{R}^2;\ y = 0\}$ est l'image réciproque du fermé $\{0\}$ de $\mathbb{R}$ par l'application $\mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto y$ qui est continue, donc est un fermé de $\mathbb{R}^2$. De même l'ensemble $B = \{(x, y) \in \mathbb{R}^2;\ xy = 1\}$ est l'image réciproque du fermé $\{1\}$ de $\mathbb{R}$ par l'application $\mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto xy$ qui est continue, donc est un fermé de $\mathbb{R}^2$.

Cependant $A + B$, qui est l'ensemble $\mathbb{R}^2 \setminus A$, n'est pas fermé.

Exercice 3

Soit f l'application définie sur $\mathbb{R}^2$ par $f(x, y) = e^{x^2+y^2}\Big(x^2 + xy + y^2 - \frac{1}{2}\Big)$.

1. La différence entre les deux membres de l'inégalité est $\frac{1}{2}(a^2 + 2ab + b^2) = \frac{1}{2}(a + b)^2 \geq 0$.

2. L'application f est coercive puisque, d'après la question 1., $f(x, y) \geq \frac{1}{2}e^{x^2+y^2}(x^2 + y^2 - 1)$ qui tend vers $+\infty$ quand $\|(x, y)\|$ tend vers l'infini. Comme de plus elle est continue, elle est minorée sur $\mathbb{R}^2$ et atteint sa borne inférieure par le théorème de Weierstrass. Elle n'est pas majorée sur $\mathbb{R}^2$ car elle est coercive.

3. L'application f est un produit de fonctions usuelles de classe C^1 sur $\mathbb{R}^2$ donc est C^1 sur $\mathbb{R}^2$. De plus

$$\frac{\partial f}{\partial x}(x, y) = e^{x^2+y^2}\Big(2x(x^2 + xy + y^2 - \frac{1}{2}) + 2x + y\Big)$$

$$\frac{\partial f}{\partial y}(x,y) = e^{x^2+y^2}\Big(2y(x^2+xy+y^2-\frac{1}{2})+x+2y\Big).$$

4.a. On a $f(x,y) \geq f(x_0,y_0)$ pour tout $(x,y) \in \mathbb{R}^2$ donc en particulier $f(x,y_0) \geq f(x_0,y_0)$ pour tout $x \in \mathbb{R}$. Autrement dit l'application partielle $x \mapsto f(x,y_0)$ admet un minimum en x_0 et sa dérivée en x_0 est donc nulle, c'est-à-dire $\dfrac{\partial f}{\partial x}(x_0,y_0) = 0$.

5. a. Comme la fonction exponentielle ne prend que des valeurs strictement positives, le système d'équations

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) & = 0 \\ \dfrac{\partial f}{\partial y}(x,y) & = 0 \end{cases}$$

équivaut à

$$\begin{cases} 2x(x^2+xy+y^2-\frac{1}{2})+2x+y & = 0 \qquad (1) \\ 2y(x^2+xy+y^2-\frac{1}{2})+x+2y & = 0 \qquad (2) \end{cases}$$

La différence de ces deux équations donne $(x-y)(x^2+xy+y^2) = 0$ qui est équivalent à $x = y$ (à l'aide de la question 1).

De (1) on déduit alors que $x(3x^2+1) = 0$, qui équivaut à $x = 0$.

Ainsi la seule solution du système est $x = y = 0$.

b. D'après la question 2 ; la fonction f admet au moins un point de minimum sur $\mathbb{R}^2$.

D'après les questions 4. et 5., si (x_0,y_0) est un point de minimum de f sur $\mathbb{R}^2$, alors nécessairement $(x_0,y_0) = (0,0)$.

Par conséquent $(0,0)$ est l'unique point de minimum de f sur $\mathbb{R}^2$, avec $f(0,0) = -\frac{1}{2}$.

6. a. L'ensemble $C(-1/2)$ est réduit au point $(0,0)$. En effet

- d'après la question 5.b. $f(0,0) = -1/2$ donc $(0,0) \in C(-1/2)$;

- de plus si $(x,y) \in C(-1/2)$ alors $f(x,y) = f(0,0)$ donc (x,y) est également un point de minimum de f sur $\mathbb{R}^2$. Or le point $(0,0)$ est l'unique point de minimum, donc $(x,y) = (0,0)$.

b. $C(k)$ est vide si et seulement si $k < -1/2$. En effet :

- le point $(0,0)$ est un point de minimum de f sur $\mathbb{R}^2$, donc $f(x,y) \geq f(0,0) = -1/2$ pour tout $(x,y) \in \mathbb{R}^2$. Autrement dit $C(k)$ est vide pour tout $k < -1/2$;

- de plus la fonction $f(x,0) = e^{x^2}(x^2 - \frac{1}{2})$ pour $x \in \mathbb{R}$ tend vers $+\infty$ quand x tend vers $+\infty$, et vérifie $f(0,0) = -1/2$ donc le théorème des valeurs intermédiaires appliqué à la fonction continue $x \mapsto f(x,0)$ assure qu'il existe $(x,0)$ dans $C(k)$ si $k \geq -1/2$: ainsi $C(k)$ est non vide pour tout $k \geq -1/2$.

c. Soit $k \in \mathbb{R}$ fixé. L'ensemble $C(k)$ est fermé comme image réciproque du fermé $\{k\}$ de $\mathbb{R}$ par l'application continue $f : \mathbb{R}^2 \to \mathbb{R}$.

De plus f est coercive, donc il existe une constante R telle que $f(x,y) \geq k+1$ pour tout (x,y) à l'extérieur de la boule de centre $(0,0)$ et de rayon R (pour une certaine norme). En particulier l'ensemble $C(k)$ est nécessairement inclus dans cette boule, et est donc borné.

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Contrôle continu du 27 mai 2008

Le sujet comporte 2 pages. L'épreuve dure 1 heure 15. Les documents, calculatrices et téléphones portables sont interdits.

Question de cours

Soit f une fonction différentiable dans $\mathbb{R}^n$, à valeurs dans $\mathbb{R}$. Montrer que la fonction f est convexe sur $\mathbb{R}^n$ si et seulement si $f(y) - f(x) \geq df(x)\,(y-x)$ pour tous x et $y \in \mathbb{R}^n$.

Exercice 1

Soit f la fonction définie par

$$f(x,y) = x^4 y - y^2 - y\,\ln(x+1) + 2.$$

1. Montrer que son ensemble de définition D est un sous-ensemble ouvert et convexe de $\mathbb{R}^2$.
2. Montrer que f est de classe $\mathcal{C}^2$ sur D et calculer ses dérivées partielles premières et secondes.
3. Écrire la formule de Taylor-Young de f à l'ordre 2 au point $(0,0)$.
4. Etudier la convexité de la fonction f sur D.
5. Montrer que f n'est ni majorée, ni minorée sur D.

Exercice 2

On dit qu'une fonction f définie sur un sous-ensemble convexe U de $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$ est quasi-concave si pour tout réel k l'ensemble

$$E_k = \{x \in U; f(x) \geq k\}$$

est convexe.
1. Montrer que la fonction $f : \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto -x^2 - y^2$ est quasi-concave mais que $-f$ ne l'est pas.
2. Montrer qu'une fonction concave sur un sous-ensemble convexe de $\mathbb{R}^n$ est quasi-concave.
3. Montrer qu'une fonction f définie sur un sous-ensemble convexe U de $\mathbb{R}^n$ est quasi-concave si et seulement si

$$f(tx + (1-t)y) \geq min(f(x), f(y))$$

pour tous $x, y \in U, t \in [0,1]$.

4. Soit f une fonction quasi-concave sur un sous-ensemble convexe U de $\mathbb{R}^n$ et g une fonction croissante sur un intervalle I de $\mathbb{R}$, avec $f(U) \subset I$. Déduire de la question 3. que $g \circ f$ est quasi-concave sur U.

5. Montrer que la fonction $h : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto e^{-x^2-y^2}$ est quasi-concave mais pas concave.

Exercice 3

Soit f une fonction définie sur un sous-ensemble convexe U de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$. Montrer que la fonction f est convexe sur U si et seulement si son épigraphe

$$Epi(f) = \{(x, k) \in U \times \mathbb{R}; f(x) \leq k\}$$

est un sous-ensemble convexe de $\mathbb{R}^{n+1}$.

Corrigé succinct du contrôle continu du 27 mai 2008

Exercice 1

Soit f la fonction définie par

$$f(x,y) = x^4y - y^2 - y\,\ln(x+1) + 2.$$

1. $f(x,y)$ est défini si et seulement si $x > -1$, donc l'ensemble de définition de f est le demi-plan ouvert $\{x, y \in \mathbb{R}^2; x > -1\}$ qui est un sous-ensemble ouvert et convexe de $\mathbb{R}^2$.

2. La fonction $x \mapsto \ln(x+1)$ est de classe C^2 sur $]-1,+\infty[$, donc $(x,y) \mapsto \ln(x+1)$ puis f est de classe C^2 sur $]-1,+\infty[\times\mathbb{R} = D$. Les dérivées partielles premières et secondes de f sont

$$\frac{\partial f}{\partial x}(x,y) = 4x^3y - \frac{y}{x+1}, \frac{\partial f}{\partial y}(x,y) = x^4 - 2y - ln(x+1),$$

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 12x^2y + \frac{y}{(x+1)^2}, \frac{\partial^2 f}{\partial x\partial y}(x,y) = \frac{\partial^2 f}{\partial y\partial x}(x,y) = 4x^3 - \frac{1}{x+1}, \frac{\partial^2 f}{\partial y^2}(x,y) = -2.$$

3. La fonction f est deux fois différentiable en $(0,0)$, et ses dérivées partielles premières et secondes en $(0,0)$ sont

$$\frac{\partial f}{\partial x}(0,0) = 0, \frac{\partial f}{\partial y}(0,0) = 0, \frac{\partial^2 f}{\partial x^2}(0,0) = 0, \frac{\partial^2 f}{\partial x\partial y}(0,0) = \frac{\partial^2 f}{\partial y\partial x}(0,0) = -1, \frac{\partial^2 f}{\partial y^2}(0,0) = -2.$$

La formule de Taylor-Young de f à l'ordre 2 au point $(0,0)$ s'écrit donc

$$f(x,y) = 2 - xy - y^2 + (x^2+y^2)\varepsilon(x,y)$$

où ε est une fonction définie sur un voisinage de $(0,0)$, continue et nulle en $(0,0)$.

4. Le déterminant de la matrice hessienne en $(0,0)$ est

$$\frac{\partial^2 f}{\partial x^2}(0,0).\frac{\partial^2 f}{\partial y^2}(0,0) - \Big(\frac{\partial^2 f}{\partial x\partial y}(0,0)\Big)^2 = -1 < 0$$

donc la la matrice hessienne $Hf(0,0)$ n'est ni semi-définie positive ni semi-défine négative et donc la fonction f n'est ni convexe ni concave sur D.

5. La fonction f n'est pas majorée sur D puisque par exemple $f(x,1) = x^4 - \ln(x+1) + 1$ tend vers $+\infty$ quand x tend vers $+\infty$ (avec $(x,1) \in D$). Elle n'est pas minorée puisque par exemple $f(0,y) = 2 - y^2$ tend vers $-\infty$ quand y tend vers $+\infty$ (avec $(0,y) \in D$).

Exercice 2

1. Montrons que la fonction $f : \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto -x^2 - y^2$ est quasi-concave. Son ensemble de définition $\mathbb{R}^2$ est convexe, et l'ensemble

$$\{(x,y) \in \mathbb{R}^2; f(x,y) \geq k\} = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 \leq -k\}$$

est l'ensemble vide si $k > 0$, le point $(0,0)$ si $k = 0$ et la boule de centre $(0,0)$ et de rayon $\sqrt{-k}$ si $k < 0$. Dans tous les cas il est convexe, et donc f est quasi-concave.

Au contraire $-f$ n'est pas quasi-concave. En effet l'ensemble

$$\{(x,y) \in \mathbb{R}^2; -f(x,y) \geq k\} = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 \geq k\}$$

est l'extérieur de la boule de centre $(0,0)$ et de rayon $\sqrt{k}$ si $k > 0$, qui n'est pas convexe.

2. Soit f une fonction concave sur un convexe U de $\mathbb{R}^n$. Soit $k \in \mathbb{R}$ fixé et montrons que l'ensemble E_k est convexe. Soit pour cela $x, y \in E_k$ et $t \in [0,1]$. Alors

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \geq tk + (1-t)k = k$$

puisque f est concave et puisque $x, y \in E_k$ respectivement. Par conséquent $tx + (1-t)y \in E_k$, qui est donc convexe, si bien que f est quasi-concave.

3. Supposons que f soit quasi-concave sur le convexe U, et soit x, y fixés dans U et $t \in [0,1]$. Appliquons la définition de la quasi-concavité de f à $k = min(f(x), f(y))$. On a $f(x) \geq k, f(y) \geq k$, c'est-à-dire $x, y \in E_k$, qui est convexe par hypothèse sur f, donc $tx + (1-t)y \in E_k$: autrement dit $f(tx + (1-t)y) \geq k$. Or $k = min(f(x), f(y))$, et donc $f(tx + (1-t)y) \geq min(f(x), f(y))$.

Inversement supposons que la fonction f vérifie

$$f(tx + (1-t)y) \geq min(f(x), f(y))$$

pour tous $x, y \in U, t \in [0,1]$.

Soit $k \in \mathbb{R}$ fixé et montrons que E_k est convexe. Soit pour cela $x, y \in E_k$ et $t \in [0,1]$. Alors $f(x) \geq k, f(y) \geq k$, et donc $min(f(x), f(y)) \geq k$. Par conséquent $f(tx + (1-t)y) \geq min(f(x), f(y)) \geq k$, ce qui signifie que $tx + (1-t)y \in E_k$, qui est donc convexe, si bien que f est quasi-concave.

4. D'après la question 3 il suffit de montrer que

$$(g \circ f)(tx + (1-t)y) \geq min((g \circ f)(x), (g \circ f)(y))$$

pour tous $x, y \in U, t \in [0,1]$.

Or f est quasi-concave, donc d'après la question 3

$$f(tx + (1-t)y) \geq min(f(x), f(y))$$

et g est croissante, donc

$$(g \circ f)(tx + (1-t)y) = g[f(x)] \geq g\big[min(f(x), f(y))\big] = min((g \circ f)(x), (g \circ f)(y)).$$

5. La fonction $f : \mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto -x^2 - y^2$ est quasi-concave d'après la question 1 et la fonction exponentielle est croissante sur $\mathbb{R}$, donc par composition h est quasi-concave sur $\mathbb{R}^2$ d'après la question 4.

Cependant la fonction h n'est pas concave sur $\mathbb{R}^2$: en effet le déterminant de sa matrice hessienne en (x,y) est

$$e^{-2(x^2+y^2)}(4 - 8x^2 - 8y^2)$$

qui peut être strictement négatif, par exemple pour $(x, y) = (1, 1)$. (La fonction h n'est donc pas non plus convexe sur $\mathbb{R}^2$.)

Exercice 3

Soit f une fonction convexe sur le convexe U et montrons que son épigraphe $Epi(f)$ est convexe. Soit pour cela $(x, k), (y, l) \in Epi(f)$ et $t \in [0, 1]$. Alors $f(x) \leq k$ et $f(y) \leq l$, et donc

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \leq tk + (1-t)l.$$

Autrement dit $(tx + (1-t)y, tk + (1-t)l) \in Epi(f)$, c'est-à-dire $t(x, k) + (1-t)(y, l) \in Epi(f)$: ainsi $Epi(f)$ est bien convexe.

Réciproquement supposons que $Epi(f)$ soit convexe, et soit $x, y \in U, t \in [0, 1]$.

Alors $(x, f(x)), (y, f(y)) \in Epi(f)$ qui est convexe, donc $t(x, f(x)) + (1-t)(y, f(y)) \in Epi(f)$: autrement dit $(tx + (1-t)y, tf(x) + (1-t)f(y)) \in Epi(f)$, c'est-à-dire

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

ce qui assure que f est convexe.

Université Paris Dauphine
DUMI2E 2e année
Calcul différentiel

Examen du 13 juin 2008

Le sujet comporte 2 pages. L'épreuve dure 2 heures. Les documents, calculatrices et téléphones portables sont interdits.

Problème 1

Soit f la fonction définie sur $\mathbb{R}^2$ par

$$f(x,y) = \begin{cases} xy\ln(x^2+y^2) & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

1. Montrer que la fonction f est continue sur $\mathbb{R}^2$.

2. Déterminer si la fonction f est majorée, minorée, coercive sur $\mathbb{R}^2$.

3. a. Montrer que f est de classe C^2 sur $\mathbb{R}^2 \setminus (0,0)$ et calculer ses dérivées partielles premières et secondes. Déterminer la matrice hessienne de f aux points $(0,a),(a,0),(a,a)$ et $(a,-a)$ pour $a \neq 0$.

b. Montrer que la fonction f admet des dérivées partielles premières en $(0,0)$ et qu'elle est de classe C^1 sur $\mathbb{R}^2$.

4. Montrer que le point $(0,0)$ est un point col de f en menant une étude locale directe.

5. Recherche des extrema de f sur $\mathbb{R}^2 \setminus \{(0,0)\}$:

a. Ecrire le système d'équations vérifié par les points critiques $(x,y) \neq (0,0)$ de f.

b. Déterminer les points critiques (x,y) de f tels que $x=0, y\neq 0$ et déterminer si ce sont des extrema de f dans $\mathbb{R}^2 \setminus (0,0)$.

c. Déterminer les points critiques (x,y) de f tels que $x\neq 0, y=0$ et déterminer si ce sont des extrema de f dans $\mathbb{R}^2 \setminus (0,0)$.

d. Déterminer les points critiques (x,y) de f tels que $x\neq 0, y\neq 0$ et déterminer si ce sont des extrema de f dans $\mathbb{R}^2 \setminus (0,0)$.

6. Conclure quant aux extrema de f dans $\mathbb{R}^2$.

7. Déterminer les éventuels extrema globaux de f sur l'ensemble

$$B = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}.$$

8. Déterminer les éventuels extrema globaux de f sur l'ensemble

$$U = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 = 2\}.$$

On pourra optimiser la fonction $g(x,y) = xy\ln(2)$.

9. Déterminer les éventuels extrema globaux de f sur l'ensemble

$$V = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 \leq 2\}.$$

Problème 2

On munit $\mathbb{R}^n$ du produit scalaire euclidien noté $< \cdot, \cdot >$ et de la norme associée notée $\|\cdot\|$. Soit α un nombre réel > 0 et f une fonction de classe C^1 sur $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ telle que

$$f(y) - f(x) \geq < \nabla f(x), \, y - x > + \frac{\alpha}{2}\|x - y\|^2 \quad \text{pour tous } x \text{ et } y \in \mathbb{R}^n.$$

1. a. Montrer qu'un point a de $\mathbb{R}^n$ est un point de minimum global de f dans $\mathbb{R}^n$ si et seulement si $df(a) = 0$.

En déduire qu'il existe au plus un point de minimum global de f dans $\mathbb{R}^n$.

b. Montrer que la fonction f est coercive dans $\mathbb{R}^n$. En déduire qu'il existe au moins un point de minimum global de f dans $\mathbb{R}^n$.

On notera a l'unique point de minimum global de f dans $\mathbb{R}^n$.

2. Soit x un point de $\mathbb{R}^n$ distinct de a. On note φ_x la fonction de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$\varphi_x(t) = f(x + t\,\nabla f(x)), \quad t \in \mathbb{R}.$$

a. Montrer que la fonction φ_x est de classe C^1 dans $\mathbb{R}$ et expliciter sa dérivée.

b. Montrer que la fonction φ_x vérifie

$$\varphi_x(t) - \varphi_x(s) \geq \varphi_x'(s)(t - s) + \frac{\alpha}{2}\|\nabla f(x)\|^2|t - s|^2 \quad \text{pour tous } s, \, t \in \mathbb{R}.$$

c. En déduire que la fonction φ_x admet un unique point de minimum global sur $\mathbb{R}$.

On notera $t(x)$ l'unique point de minimum global de φ_x dans $\mathbb{R}$.

3. Soit x_0 un point fixé de $\mathbb{R}^n$ et $(x_k)_{k\geq 0}$ la suite de points de $\mathbb{R}^n$ définis de proche en proche en posant $x_{k+1} = a$ si $x_k = a$ et $x_{k+1} = x_k + t(x_k)\,\nabla f(x_k)$ si $x_k \neq a$.

On se propose de montrer que la suite $(x_k)_{k\geq 0}$ converge vers a dans $\mathbb{R}^n$.

Que peut-on dire si $x_k = a$ pour un entier k ?

On suppose dans la suite que $x_k \neq a$ pour tout entier k.

a. Montrer que pour tout entier $k \geq 0$ on a

$$< \nabla f(x_{k+1}), \nabla f(x_k) >= 0,$$

b. En déduire les deux inégalités

$$f(x_k) - f(x_{k+1}) \geq \frac{\alpha}{2}\|x_{k+1} - x_k\|^2,$$

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \geq \|\nabla f(x_k)\|^2.$$

c. Montrer que la suite $(f(x_k))_{k\geq 0}$ est décroissante. En déduire que la suite $(f(x_k))_{k\geq 0}$ est convergente dans $\mathbb{R}$.

d. En déduire que la suite $(x_{k+1} - x_k)_{k\geq 0}$ tend vers 0 dans $\mathbb{R}^n$ et que la suite $(x_k)_{k\geq 0}$ est contenue dans une boule fermée bornée de $\mathbb{R}^n$.

e. En déduire que la suite $(\nabla f(x_{k+1}) - \nabla f(x_k))_{k\geq 0}$ tend vers 0 dans $\mathbb{R}^n$ dans le cas où la fonction f est de plus de classe C^2 dans $\mathbb{R}^n$. On admettra ce résultat dans le cas général.

En déduire que la suite $(\nabla f(x_k))_{k\geq 0}$ tend vers 0 dans $\mathbb{R}^n$.

f. En déduire que toute suite extraite de la suite $(x_k)_k$ contient une sous-suite qui converge vers a dans $\mathbb{R}^n$. En déduire que toute la suite $(x_k)_k$ converge vers a dans $\mathbb{R}^n$.

Corrigé de l'examen du 13 juin 2008 - Problème 2

1. a - Si a est un point de minimum global de f dans l'ouvert $\mathbb{R}^n$, alors nécessairement $df(a) = 0$.

Inversement si $df(a) = 0$, alors on a pour tout $y \in \mathbb{R}^n$

$$f(y) - f(a) \geq \frac{\alpha}{2}\|y - a\|^2 \geq 0$$

et donc a est un point de minimum global dans $\mathbb{R}^n$.

(ce résultat est général pour les fonctions convexes dans l'ouvert $\mathbb{R}^n$ et en particulier pour la fonction f qui est convexe dans l'ouvert $\mathbb{R}^n$)

L'inégalité précédente montre que si a est un point de minimum global de f dans $\mathbb{R}^n$, alors pour tout $y \in \mathbb{R}^n$ avec $y \neq a$ on a

$$f(y) - f(a) > 0$$

et donc a est un unique point de minimum global dans $\mathbb{R}^n$.

(ce résultat est général pour les fonctions strictement convexes dans l'ouvert $\mathbb{R}^n$ et en particulier pour la fonction f qui est strictement convexe dans l'ouvert $\mathbb{R}^n$)

b - De la minoration de $f(y) - f(x)$ appliquée par exemple au couple $(0, y)$, on déduit que pour tout $y \in \mathbb{R}^n$ on a

$$f(y) \geq f(0) + < \nabla f(0)\,, y > + \frac{\alpha}{2}\|y\|^2 \geq f(0) - \|df(0)\|\,\|y\| + \frac{\alpha}{2}\|y\|^2.$$

Comme pour a, b et c réels avec $a > 0$, le trinôme $az^2 + bz + c$ tend vers $+\infty$ quand z tend vers $+\infty$ dans $\mathbb{R}$, il en résulte que $f(y)$ tend vers $+\infty$ quand $\|y\|$ tend vers $+\infty$, autrement dit la fonction f est coercive dans $\mathbb{R}^n$.

Ainsi comme la fonction f est coercive et continue dans $\mathbb{R}^n$, le théorème de Weierstrass assure que la fonction f a au moins un point a de minimum global dans $\mathbb{R}^n$.

2. a - La fonction (affine) g_x de $\mathbb{R}$ dans $\mathbb{R}^n$ définie pour $t \in \mathbb{R}$ par

$$g_x(t) = x + t\,\nabla f(x)$$

est de classe C^1 dans $\mathbb{R}$, dont la dérivée (constante) est donnée pour $t \in \mathbb{R}$ par

$$g_x'(t) = \nabla f(x).$$

La fonction φ_x étant la composée des fonctions f et g_x qui sont de classe C^1 est donc aussi de classe C^1 dans $\mathbb{R}$ dont la dérivée est donnée pour $t \in \mathbb{R}$ par

$$\varphi_x'(t) = df(g(t))\,(g'(t)) = < \nabla f(x + t\nabla f(x)), \nabla f(x) > .$$

b - Pour $t > s$ on a

$$\begin{aligned}\varphi_x(t)-\varphi_x(s) &= f(x+t\nabla f(x))-f(x+s\nabla f(x))\\ &\geq \; < \nabla f(x+s\nabla f(x)), (x+t\nabla f(x))-(x+s\nabla f(x)) >\\ &\quad +\frac{\alpha}{2}\|x+t\nabla f(x))-(x+s\nabla f(x))\|^2\\ &= \varphi_x'(s)(t-s)+\frac{\alpha}{2}\|\nabla f(x)\|^2|t-s|^2.\end{aligned}$$

c - Par conséquent d'après 1 - b appliquée à la fonction φ_x, on en déduit que la fonction φ_x admet un minimum global sur $\mathbb{R}$, atteint en un unique point qui sera noté $t(x)$.

3. Si $x_{k_0} = a$ pour un certain k_0 alors par construction $x_k = a$ pour tout $k \geq k_0$ et donc toute la suite $(x_k)_k$ converge (trivialement) vers a.

On suppose maintenant que $x_k \neq a$ pour pour tout $k \geq 0$.

a - D'après ce qui précède on rappelle que pour tout $x \neq a$, on a d'une part pour tout $t \in \mathbb{R}$

$$\varphi_x'(t) =< \nabla f(x+t\nabla f(x)), \nabla f(x) >$$

et on a d'autre part

$$\varphi_x'(t(x)) = 0.$$

Donc pour $x = x_k$ et $t = t(x_k)$, ces deux égalités donnent

$$< \nabla f(x_{k+1}), \nabla f(x_k) >= \varphi_{x_k}'(t(x_k)) = 0.$$

b - Par conséquent par l'hypothèse sur f on a

$$\begin{aligned}f(x_k)-f(x_{k+1}) &\geq \; < \nabla f(x_{k+1}), x_k - x_{k+1} > +\frac{\alpha}{2}\|x_{k+1}-x_k\|^2\\ &= -< \nabla f(x_{k+1}), t(x_k)\,\nabla f(x_k) > +\frac{\alpha}{2}\|x_{k+1}-x_k\|^2\\ &= -t(x_k)\; < \nabla f(x_{k+1}), \nabla f(x_k)) > +\frac{\alpha}{2}\|x_{k+1}-x_k\|^2 = \frac{\alpha}{2}\|x_{k+1}-x_k\|^2,\end{aligned}$$

et en développant le carré scalaire on a

$$\begin{aligned}\|\nabla f(x_{k+1})-\nabla f(x_k)\|^2 &= \|\nabla f(x_{k+1})\|^2 - 2 < \nabla f(x_{k+1}), \nabla f(x_k) > + \|\nabla f(x_k)\|^2\\ &= \|\nabla f(x_{k+1})\|^2 + \|\nabla f(x_k)\|^2 \geq \|\nabla f(x_k)\|^2.\end{aligned}$$

c - On peut montrer que la suite $(f(x_k))_k$ est décroissante de deux manières. D'une part par l'inégalité précédente, il est immédiat que la suite $(f(x_k))_k$ est décroissante.

D'autre part par définition

$$f(x_{k+1}) = f(x_k + t(x_k)\,\nabla f(x_k)) = \varphi_{x_k}(t(x_k))$$

et $t(x_k)$ étant le point de minimum de φ_{x_k} on a en particulier

$$\varphi_{x_k}(t(x_k)) \leq \varphi_{x_k}(0) = f(x_k).$$

Ainsi la suite $(f(x_k))_k$ est décroissante.

Comme de plus f est minorée dans $\mathbb{R}^n$ on en déduit que cette suite converge dans $\mathbb{R}$.

d - Comme la suite $(f(x_k))_k$ converge dans $\mathbb{R}$, la suite $(f(x_k) - f(x_{k+1})_k$ converge vers 0 dans $\mathbb{R}$. Par conséquent l'inégalité de b permet d'en déduire que la suite $(x_k - x_{k+1})_k$ converge vers 0.

Si la suite $(x_k)_k$ n'est pas bornée dans $\mathbb{R}^n$, il existe une sous-suite $(x_{k_j})_j$ telle que la suite $(\|x_{k_j}\|)_j$ tende vers $+\infty$. Comme la fonction f est coercive alors la suite $(f(x_{k_j}))_j$ tend vers $+\infty$. Ce qui est contraire au fait que la suite $(f(x_k))_k$ est convergente dans $\mathbb{R}$, donc bornée. Ainsi la suite $(x_k)_k$ est bornée dans $\mathbb{R}^n$, donc contenue dans une boule fermée bornée B de $\mathbb{R}^n$.

e - On suppose f de classe C^2 dans $\mathbb{R}^n$. Par l'inégalité des accroissements finis appliquée à la fonction ∇f de classe C^1 dans $\mathbb{R}^n$ valeurs dans $\mathbb{R}^n$ et à l'intervalle fermé d'extrémités x_{k+1} et x_k, il existe $\theta \in]0,1[$ tel que

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \|d(\nabla f)(x_k + \theta(x_{k+1} - x_k))(x_{k+1} - x_k)\|.$$

Or d'une part la suite $(x_k)_k$ étant contenue dans une boule fermée bornée B de $\mathbb{R}^n$. l'intervalle fermé d'extrémités x_{k+1} et x_k est contenu dans B. D'autre part la fonction f étant de classe C^2 dans $\mathbb{R}^n$, ses dérivées partielles d'ordre deux sont continues dans $\mathbb{R}^n$, donc bornées dans la boule fermée bornée B. Par suite il existe une constante $C > 0$ telle que pour tout k on ait

$$\|d(\nabla f)(x_k + \theta(x_{k+1} - x_k))(x_{k+1} - x_k)\| \leq C\, \|x_{k+1} - x_k\|$$

et donc

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \, C\, \|x_{k+1} - x_k\|.$$

Comme la suite $(x_{k+1} - x_k)_k$ tend vers 0 dans $\mathbb{R}^n$, il en résulte que la suite $(\nabla f(x_{k+1}) - \nabla f(x_k))_k$ tend vers 0 dans $\mathbb{R}^n$.

Enfin d'après b on a pour tout k

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 \geq \|\nabla f(x_k)\|^2,$$

donc

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \geq \|\nabla f(x_k)\| \geq 0$$

et par conséquent il en résulte que la suite $(\nabla f(x_k))_k$ tend aussi vers 0 dans $\mathbb{R}^n$.

f - Soit $(x_{k_j})_j$ une sous-suite de la suite $(x_k)_k$. Comme elle est contenue dans le fermé borné B, elle contient une sous-suite que l'on notera encore $(x_{k_j})_j$ pour simplifier, qui est convergente vers un certain point b dans B. Comme la fonction ∇f est continue dans $\mathbb{R}^n$, la sous-suite associée $(\nabla f(x_{k_j}))_j$ converge vers $\nabla f(b)$. Or d'après e cette sous-suite converge vers 0. Donc d'après l'unicité de la limite c'est que nécessairement $\nabla f(b) = 0$ et comme on l'a vu en 1-a, nécessairement b est alors égal à a.

Ainsi de toute suite extraite de la suite $(x_k)_k$ on peut extraire une sous-suite qui converge vers la même limite a. Par conséquent par l'absurde on en déduit que toute la suite $(x_k)_k$ converge vers a.

Université Paris-Dauphine
DUMI2E 2e année
Calcul différentiel

Examen de septembre 2008 - Deuxième session

Le sujet comporte 2 pages. L'épreuve dure 2 heures. Les documents, calculatrices et téléphones portables sont interdits.

Exercice 1

Soit $\|\cdot\|$ une norme sur $\mathbb{R}^n$, a un point de $\mathbb{R}^n$ et B une partie de $\mathbb{R}^n$, et notons

$$d(a, B) = \inf_{b \in B} \|a - b\|.$$

1. Montrer que pour tout entier n il existe $b_n \in B$ tel que $d(a, B) \geq \|a - b_n\| - 1/n$.

2. Si B est compact, montrer qu'il existe $b \in B$ tel que $d(a, B) = \|a - b\|$.

3. Si B est fermé, montrer qu'il existe $b \in B$ tel que $d(a, B) = \|a - b\|$.

4. Si B est borné, montrer qu'il existe $b \in \mathbb{R}^n$ tel que $d(a, B) = \|a - b\|$.

5. Dans $\mathbb{R}^2$, soit a le point $(0, 1)$ et B l'ensemble $\{1\} \times [-1, +1[$. Déterminer l'ensemble des points $b \in B$ tels que $d(a, B) = \|a - b\|$ dans les deux cas suivants :

a. si $\mathbb{R}^2$ est muni de la norme euclidienne $\|(x, y)\|_2 = (x^2 + y^2)^{1/2}$ pour $(x, y) \in \mathbb{R}^2$;

b. si $\mathbb{R}^2$ est muni de la norme infinie $\|(x, y)\|_\infty = \max(|x|, |y|)$ pour $(x, y) \in \mathbb{R}^2$.

Exercice 2.

Soit f une fonction définie dans $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et soit $G(f)$ son graphe, c'est-à-dire le sous-ensemble de $\mathbb{R}^n \times \mathbb{R}$ défini par

$$G(f) = \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R};\ x \in \mathbb{R}^n\}.$$

On munit l'ensemble $\mathbb{R}^n$ d'une norme $\|\cdot\|$, l'ensemble $\mathbb{R}$ de sa valeur absolue $|\cdot|$ et l'ensemble $\mathbb{R}^n \times \mathbb{R}$ (identifié à $\mathbb{R}^{n+1}$) d'une norme associée, comme par exemple $\|(x, y)\| = \|x\| + |y|$.

1. Si la fonction f est continue dans $\mathbb{R}^n$, montrer que son graphe $G(f)$ est fermé dans $\mathbb{R}^n \times \mathbb{R}$ (on pourra utiliser les suites).

2. Si l'image $f(\mathbb{R}^n) = \{f(x) \in \mathbb{R};\ x \in \mathbb{R}^n\}$ de la fonction f est contenue dans un fermé borné de $\mathbb{R}$ et si son graphe $G(f)$ est fermé dans $\mathbb{R}^n \times \mathbb{R}$, montrer que la fonction f est continue dans $\mathbb{R}^n$ (on pourra raisonner par l'absurde et utiliser les suites).

Exercice 3.

Soit $D = \{(x,y) \in \mathbb{R}^2 ; x > 0, y > 0\}$, α et β deux réels strictement positifs et f la fonction définie sur D par $f(x,y) = x^\alpha y^\beta$.

1. Si k est un réel on note $C(k) = \{(x,y) \in D; f(x,y) = k\}$ la courbe de niveau k de la fonction f.

Montrer que pour tout réel $k > 0$ l'ensemble $C(k)$ est le graphe d'une fonction convexe dune variable réelle.

Déterminer $C(k)$ pour $k \leq 0$.

2. a. Montrer que la fonction f est de classe C^2 sur D.

b. Déterminer une condition nécessaire et suffisante sur α et β pour que f soit concave sur D.

c. Montrer qu'il n'existe pas de α et β tels que f soit convexe sur D.

3. Soit α et β deux réels strictement positifs quelconques. Montrer que la fonction f peut s'écrire sous la forme $h \circ g$ où g est une fonction concave sur D à valeurs dans $]0, +\infty[$ et h est une fonction strictement croissante de $]0, +\infty[$ sur $0, +\infty[$.

4. Pour tout réel k, montrer que l'ensemble $\{(x,y) \in D; f(x,y) \geq k\}$ est un sous-ensemble convexe de $\mathbb{R}^2$ (on pourra utiliser le résultat de la question 3).

5. Ecrire la formule de Taylor-Young à l'ordre 2 de f au point $(1,1)$ et en déduire une valeur approchée de la quantité $\sqrt{1,2 \times 0,9}$ (on choisira α et β).

Exercice 4.

Soit f la fonction définie sur $\mathbb{R}^2$ par

$$f(x,y) = \begin{cases} \dfrac{x^2y^2}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases}$$

1. Montrer que la fonction f est continue sur $\mathbb{R}^2$.

2. Déterminer si la fonction f est majorée, minorée, coercive sur $\mathbb{R}^2$.

3. a. Montrer que f est de classe C^1 sur $\mathbb{R}^2 \setminus (0,0)$ et calculer ses dérivées partielles premières.

b. Montrer que la fonction f admet des dérivées partielles premières en $(0,0)$ et qu'elle est de classe C^1 sur $\mathbb{R}^2$.

4. Déterminer les extrema de la fonction f sur $\mathbb{R}^2$ et préciser leur nature (maximum, minimum, global, local).

5. Déterminer les extrema de la fonction f sur $B = \{(x,y) \in \mathbb{R}^2 ; x^2 + y^2 < 1\}$ et préciser leur nature (maximum, minimum, global, local).

6. Déterminer les extrema de la fonction f sur $U = \{(x,y) \in \mathbb{R}^2 ; x^2 + y^2 = 1\}$ et préciser leur nature (maximum, minimum, global, local).

7. Déterminer les extrema globaux de la fonction f sur $V = \{(x,y) \in \mathbb{R}^2 ; x^2 + y^2 \leq 1.\}$

Table des matières