Université Claude Bernard Lyon 1 **Licence Mathématiques et économie**

Analyse pour l'économie 1

Lorenzo Brandolese

1 Intégrales impropres

...

2 Séries numériques

...

3 Topologie dans $\mathbb{R}^n$

3.1 Distances et normes

Définition 3.1 (Distance et espace métrique)**.** *Soit X un ensemble non vide. Une* distance *(ou* métrique*) sur X est une application $d\colon X \times X \to \mathbb{R}^+$, telle que, pour tout x, y et $z \in X$,*

i) $d(x,y) = 0 \iff x = y$,

ii) $d(x,y) = d(y,x)$,

iii) $d(x,y) \le d(x,z) + d(z,y)$ *(Inégalité triangulaire).*

Un espace métrique *est un couple (X,d) formé par un ensemble et une distance (mais on écrira souvent seulement X pour simplifier). Les éléments de X s'appellent* points.

Définition 3.2 (Norme et espace normé)**.** *Soit E un espace vectoriel réel. Une norme sur E est une application $\|\cdot\|\colon E \to \mathbb{R}^+$ telle que, pour tout $x, y \in E$ et tout $\lambda \in R$ on a:*

j) $\|x\| = 0 \iff x = 0_E$,

jj) $\|\lambda x\| = |\lambda|\,\|x\|$ *(homogénéité de la norme).*

jjj) $\|x+y\| \le \|x\| + \|y\|$. *(Inégalité triangulaire pour les normes).*

Un espace vectoriel normé est le couple $(E, \|\cdot\|)$ formé par un espace vectoriel et une norme sur E. Les éléments de E s'appellent vecteurs *ou* points.

Exemple 3.1.

1. En prenant $E = \mathbb{R}$, on voit que la fonction "valeur absolue" $x \mapsto |x|$ verifie les propriétés j)–jjj).

2. En prenant $E = \mathbb{R}^n$, on définit la *norme-1*, notée $\|\cdot\|_1$, par $\|x\|_1 = \sum_{i=1}^n |x_i|$, pour tout $x = (x_1, \dots x_n) \in \mathbb{R}^n$.

3. En prenant $E = \mathbb{R}^n$, on définit la *norme-∞* (ou norme du max), notée $\|\cdot\|_\infty$, par $\|x\|_\infty = \max_{i=1}^n |x_i|$, pour tout $x = (x_1, \dots x_n) \in \mathbb{R}^n$.

Exemple 3.2 (Norme euclidienne)**.** En prenant $E = \mathbb{R}^n$, on définit la *norme euclidienne* (ou norme usuelle), notée $\|\cdot\|_2$, par $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$, pour tout $x = (x_1, \dots x_n) \in \mathbb{R}^n$. La vérification de l'inégalité triangulaire n'est pas immédiate: elle repose sur l'inégalité de Cauchy-Schwarz ci dessous. La norme euclidienne dans $\mathbb{R}^n$ est parfois notée simplement $\|\cdot\|$, ou encore $|\cdot|$. Par exemple, dans $\mathbb{R}^3$, $\|(1,2,3)\| = \sqrt{14}$.

Rappelons que le *produit scalaire* entre deux vecteurs de $\mathbb{R}^n$ est défini par

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

Observons que l'on a

$$\|x\|_2 = \sqrt{\langle x, x\rangle}.$$

Proposition 3.1 (Inégalité de Cauchy-Schwarz)**.** *Pour tout $x = (x_1, \dots x_n)$ et $y = (y_1, \dots y_n)$ vecteurs de $\mathbb{R}^n$ on a*

$$|\langle x, y\rangle| \le \|x\|_2 \ \|y\|_2.$$

Pour la démonstration de cette inégalité nous renvoyons au cours d'algèbre linéaire. Notons que les autres normes sur $\mathbb{R}^n$ ne vérifient aucune relation de ce type avec le produit scalaire.

Nous pouvons maintenant démontrer l'inégalité triangulaire pour la norme euclidienne: observons d'abord que

$$\langle x+y, x+y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle.$$

L'égalité ci-dessus se réécrit

$$\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2.$$

L'inégalité de Cauchy-Schwarz implique alors

$$\|x+y\|^2 \le \|x\|^2 + 2\|x\| \ \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$$

et l'inégalité triangulaire $\|x+y\| \le \|x\| + \|y\|$ en découle, en prenant terme-à-terme la racine carrée dans la dernère inégalité.

Proposition 3.2. *Dans un espace vectoriel normé $(E, \|\cdot\|)$ et $X \subset E$, on peut toujours définir une distance entre deux points de X de la manière suivante:*

$$\forall\, x, y \in X : d(x,y) = \|x - y\|.$$

En particulier tout espace vectoriel normé hérite la structure d'espace métrique.

En particulier, sur $\mathbb{R}^n$ on obtient de cette manière la *distance euclidienne* (ou distance usuelle), notée d_2, ou parfois simplement d s'il n'y a pas de risque de confusion. Plus explicitement, pour tout $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n) \in \mathbb{R}^n$,

$$d_2(x,y) = \Big(\sum_{i=1}^n |x_i - y_i|^2\Big)^{1/2}.$$

Dans $\mathbb{R}^n$, on considérera toujours la distance euclidienne, sauf indication contraire.

Définition 3.3. *On dit que deux distances d et d' sur un même ensemble X sont* (métriquement) équivalentes *(ou* quasi-isométriques*) s'il existent deux constantes réelles $0 < A < B$ telles que pour tout $x, y \in X$ on a*

$$Ad(x,y) \leq d'(x,y) \leq Bd(x,y).$$

De la même manière on peut définir, sur un espace vectoriel E la notion de *normes équivalentes*: cela signifie qu'il existe $0 < A < B$ telles que $A\|\cdot\| \leq \|\!|\cdot|\!\| \leq B\|\cdot\|$. Si deux normes sont équivalentes les distances correspondantes qu'elles induisent seront bien entendu équivalentes.

Définition 3.4 (Partie bornée)**.** *Une partie A d'un espace métrique est dite* bornée *s'il existe $M > 0$ telle que pour tout $x, y \in A$ on a $d(x,y) \leq M$.*

Définition 3.5. *Soit (X,d) un espace métrique, $x \in X$ et $r > 0$. On note*

1. *$B(x,r) = \{y \in X : d(x,y) < r\}$ (boule ouverte de centre x et rayon r).*
2. *$\overline{B}(x,r) = \{y \in X : d(x,y) \leq r\}$ (boule fermée de centre x et rayon r).*

Définition 3.6.
1. *Une partie U d'un espace métrique (X,d) est dite* ouverte dans X *si, pour tout $x \in U$ il existe $r > 0$ telle que $B(x,r) \subset U$.*
2. *Une partie U d'un espace métrique (X,d) est dite* fermée dans X *si l'ensemble complémentaire $X \backslash U$ est une partie ouverte dans X.*

Remarque 3.3.

- Dans un espace métrique (X,d), l'ensemble vide $\emptyset$ et l'espace X tout entier sont toujours *simultanément ouverts et fermés* dans X. Bien souvent, il arrive que l'ensemble vide et l'espace entier X soient les seules parties à être simultanement ouvertes et fermées : on dit alors que, par définition, (X,d) est un espace *connexe*.
- Dans $\mathbb{R}$ les intervalles de type $]a,b[$ sont des parties ouvertes. Les intervalles de la forme $]-\infty,a[$ et $]b,+\infty[$ sont aussi des parties ouvertes, comme on le vérifie l'aide de la définition.
- Les intervalles de type $[a,b]$ sont des parties fermées dans $\mathbb{R}$, au sens ci-dessus. En effet, le complémentaire de cet intervalle est la réunion $]-\infty,a[\cup]b,+\infty[$, qui est un ensemble ouvert (la réunion d'ouverts étant un ouvert, voir la proposition ci-dessus).
- L'intervalle $[a,b[$ est une partie de $\mathbb{R}$ qui n'est ni ouverte, ni fermée.

Proposition 3.3. *Dans un espace métrique, pour tout $x \in X$ et $r > 0$, les boules $B(x,r)$ sont des parties ouvertes et $\overline{B}(x,r)$ sont bien des parties fermées dans X.*

Dém. Démonstration* □

Proposition 3.4.

- *Dans un espace métrique, la réunion (finie ou infinie) de partie ouvertes est une partie ouverte. L'intersection* finie *de parties ouvertes est une partie ouverte.*
- *L'intersection (finie ou infinie) de parties fermées est une partie fermée. La réunion finie de parties fermée est une partie fermée.*

*Soit y un point arbitraire tel que $y \in B(x,r)$. Posons $r' = r - d(x,y)$. On a bien $r' > 0$. Vérifions que que si $z \in B(y,r')$ alors $z \in B(x,r)$: en effet, $d(x,z) \leq d(x,y) + d(y,z) < d(x,y) + r' = r$. Mais alors $B(y,r') \subset B(x,r)$. Cela prouve que $B(x,r)$ est un ouvert.

On montre de manière semblable que l'ensemble $\{y \in X : d(x,r) > r\}$ est un ouvert de X. Par passage au complémentaire, il en découle que $\overline{B}(x,r)$ est bien une partie fermée de X.

Dém. Démonstration* □

Définition 3.7 (Intérieur et adhérence)**.** *Si A est une partie d'un espace métrique, l'*intérieur *de A, noté $\overset{\circ}{A}$ est la réunion de tous les ouverts contenus dans A. L'*adhérence *de A, notée $\overline{A}$, est l'intersection de toutes les parties fermées contenant A.*

Exemple 3.4. Dans l'espace métrique $\mathbb{R}$, muni de la distance usuelle, si $A = [a, b[$, On a $\overset{\circ}{A}=]a, b[$ et $\overline{A} = [a, b]$.

Remarque 3.5. Si A est une partie d'un espace métrique on a par définition $\overset{\circ}{A} \subset A \subset \overline{A}$.
De plus $\overset{\circ}{A}$ est ouvert et c'est le plus grand ouvert contenu dans A. D'autre part, $\overline{A}$ est fermé, est c'est le plus petit fermé contenant A.

Un ensemble sera alors ouvert si et seulement si il coïncide avec son intérieur et férmé si et seulement s'il coïncide avec son adhérence.

Définition 3.8 (voisinage)**.** *Soit x un point d'un espace métrique X. Tout ensemble A tel que $x \in \overset{\circ}{A}$ s'appelle* voisinage de x*. On dit qu'une propriété* est satisfaite au voisinage de x *si elle est vérifiée au moins dans une boule contenant x.*

Exemple 3.6. La fonction $f : \mathbb{R} \to \mathbb{R}$ définie par $f(0) = 0$ et $f(x) = 1/x$ pour $x \neq 0$ est *bornée au voisinage de* 2. Elle n'est pas bornée au voisinage de 0. La fonction $g(x) = x - 1$ est ≥ 0 au voisinage de $x = 5/3$. Mais elle n'est pas ≥ 0 au voisinage de $x = 1$.

3.2 Suites dans un espace métrique

Une suite dans un espace métrique X est une application $\mathbb{N} \to X$. Elle est notée généralement $(x_n)_{n\in\mathbb{N}}$ ou simplement (x_n).

Définition 3.9. *Soit (x_n) une suite d'un espace métrique (X, d) et $x \in X$. On dit que la suite (x_n) converge vers x, et on écrit $x_n \to x$ ou encore $\lim_{n\to+\infty} x_n = x$ si, pour tout $\epsilon > 0$ il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$ on a $x_n \in B(x, \epsilon)$ (autrement dit, $d(x_n, x) < \epsilon$). Si la suite ne converge vers aucun point, on dit qu'elle* diverge.

Dans le cas particulier $X = \mathbb{R}$ on retrouve la définition usuelle de convergence d'une suite réelle. Dans le cas général on voit que $x_n \to x \iff d(x_n, x) \to 0$.

Proposition 3.5 (Propriétés principales des suites)**.**

- *Si (x_n) est une suite d'un espace métrique et $x_n \to x$ et $x_n \to y$, alors $x = y$. (Unicité de la limite).*
- *Toute suite convergente est bornée.*
- *Si $\lim_{n\to\infty} x_n = x$ alors toute suite extraite $(x_{n_k})_{k\in N}$ (où $(n_k)_{k\in\mathbb{N}}$ est une suite de naturels strictement croissante) vérifie $\lim_{k\to\infty} x_{n_k} = x$.*
- *Si d et d' sont deux distances équivalentes sur un même ensemble X, on a $x_n \overset{d}{\to} x$ si et seulement si $x_n \overset{d'}{\to} x$.*

*Si $(U_i)_{i\in I}$ est une famille d'ouverts (l'ensemble d'indices I pouvant être fini ou infini) et $x \in \bigcup_{i\in I} U_i$, alors il existe au moins un indice $i \in I$ tel que $x \in U_i$ et donc il existe $r > 0$ tel que $B(x, r) \subset U_i$. Mais alors $B(x, r)$ est contenue dans $\bigcup_{i\in I} U_i$ et cet ensemble est ouvert.

Si U_1 et U_2 sont ouverts et $x \in U_1 \cap U_2$, alors il existe $r_1, r_2 > 0$ tels que les boules centrées en x de rayon r_1 et r_2 sont contenues respectivement dans U_1 et dans U_2. Posons $r = \min\{r_1, r_2\}$. Alors $B(x, r) \subset U_1 \cap U_2$ ce qui montre que $U_1 \cap U_2$ est ouvert. Cet argument se généralise immédiatement à une intersection finie $U_1 \cap \ldots \cap U_n$.

Les autres affirmations se démontrent par passage au complémentaire.

En particulier, si on peut extraire d'une même suite deux suites ayant deux limites différentes, la suite de départ est nécessairement divergente.

Proposition 3.6. *Soit A une partie d'un espace métrique X et $x \in X$. On a $x \in \overline{A}$ si et seulement s'il existe une suite $(x_n) \subset A$ telle que $x_n \to A$. En particulier, un ensemble A est fermé si et seulement si:*

$$\forall\, (x_n) \subset A \text{ telle que } x_n \to x \text{ on a } x \in A.$$

Dém. Supposons $x \in \overline{A}$. Alors pour tout $n \in \mathbb{N}^*$ on peut trouver $x_n \in B(x, \frac{1}{n}) \cap A$. En effet, sinon, un aurait $B(x, \frac{1}{n}) \cap A = \emptyset$ et donc $A \subset B(x, \frac{1}{n})^c$; mais alors $\overline{A} \subset B(x, \frac{1}{n})^c$ (puisque cet ensemble est un fermé contenant A). C'est absurde puisque $x \in \overline{A}$ et $x \notin B(x, \frac{1}{n})^c$. On a $0 \leq d(x_n, x) \leq \frac{1}{n}$ et donc $x_n \to x$ par le théorème des gendarmes ; de plus la suite (x_n) est bien contenues dans A.

Réciproquement, soit $(x_n) \subset A$ et $x_n \to x$. Si, par contradiction, $x \notin \overline{A}$, alors $x \in (\overline{A})^c$, qui est ouvert. On trouve alors $r > 0$ tel que $B(x, r) \subset (\overline{A})^c$. Mais $x_n \to x$ et donc il existe $n_0 \in \mathbb{N}$ tel que $(x_n)_{n \geq n_0} \subset B(x, r) \subset (\overline{A})^c \subset A^c$. Cela contredit le fait que $(x_n) \subset A$. $\square$

Proposition 3.7 (Suites dans $\mathbb{R}^d$)**.** *Une suite (x_n) dans $\mathbb{R}^d$ converge vers $x \in \mathbb{R}^d$ si et seulement si elle converge* composante par composante. *Autrement dit, si $x = (x^1, \ldots, x^d)$ est le vecteur des composantes de x et $x_n = (x_n^1, \ldots x_n^d)$ est le vecteur des composantes de x_n, on a:*
$x_n \to x$ (pour la distance euclidienne dans $\mathbb{R}^d$) si et seulement si, pour $i = 1, \ldots, d$ on a $x_n^i \to x^i$ pour la distance usuelle de $\mathbb{R}$.

3.3 Limites de fonctions entre espaces métriques et continuité

Dans toute cette section on considère un espace métrique (X, d_X) et un espace métrique (Y, d_Y). De plus soit $x_0 \in X$ et $y_0 \in Y$. Nous aurons affaire avec des applications $f \colon D \subset X \to Y$, où D est l'ensemble de définition de la fonction f. Bien souvent, l'ensemble e définition ne sera pas explicité, mais on pourra le déterminer à l'aide de l'expression de f. Par exemple, la fonction de deux variables $f(x, y) = \frac{1}{x^2+y^2}$ est une fonction $f \colon \mathbb{R}^2 \to \mathbb{R}$, dont l'ensemble de définition est $D = \{(x, y) \in \mathbb{R}^2 \colon (x, y) \neq (0, 0)\}$.

Définition 3.10. *Soit $x_0 \in \overline{D}$ (l'adhérence dans (X, d_X) de l'ensemble de définition de f). On dit que $f(x)$ tend vers y_0 pour $x \to x_0$ dans D, et on écrit*

$$\lim_{\substack{x \to x_0 \\ x \in D}} f(x) = y_0$$

(ou, plus simplement, $\lim_{x \to x_0} f(x) = y_0$ lorsque $D = X$ ou $D = X \backslash \{x_0\}$), si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que:

$$\forall x \in D \backslash \{x_0\}, \text{ tel que } d_X(x, x_0) < \delta \quad \text{on a } \; d_Y(f(x), y_0) < \epsilon.$$

Noter que, dans cette définition, la valeur de f au point x_0 ne joue aucun rôle.

Proposition 3.8 (Limites de fonctions et suites)**.** *Soit $f \colon D \subset X \to Y$. On a $\lim_{\substack{x \to x_0 \\ x \in D}} f(x) = y_0$ si et seulement si pour toute suite $(x_n) \subset D \backslash \{x_0\}$ telle que $x_n \to x_0$, on a $f(x_n) \to y_0$.*

Dém. Démonstration* $\square$

*Supposons $\lim_{\substack{x \to x_0 \\ x \in D}} f(x) = y_0$. Soit $x_n \to x_0$. Pour tout $\epsilon > 0$ il existe $\delta > 0$ tel que si $x \in D \backslash \{x_0\}$ et $d_X(x_n, x) < \delta$ alors $d_Y(f(x_n), y_0) < \epsilon$. Mais $x_n \to x$ et donc à partir d'un certain rang n_0, $d(x_n, x_0) < \delta$ et la conclusion précédente s'applique, ce qui implique $f(x_n) \to f(x_0)$.

Réciproquement, par contraposée, si on n'a pas $\lim_{x \to x_0} f(x) = y_0$ alors il existe $\epsilon > 0$ tel que pour tout $\delta > 0$ on trouve $x \in D \backslash \{x_0\}$ tel que $0 < d(x, x_0) < \delta$ et $d_Y(f(x), y_0) \geq \epsilon$. En appliquant ceci avec $\delta = 1/n$, $n = 1, 2, \ldots$ on construit une suite $(x_n) \subset D \backslash \{x_0\}$ telle que la suite $f(x_n)$ ne converge pas vers y_0. $\square$

Définition 3.11. *On dit que f est continue en x_0 si, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que:*

$$d_X(x, x_0) < \delta \Rightarrow d_Y(f(x), f(x_0)) < \epsilon.$$

On dit qu'une fonction est continue en A (où $A \subset X$) si elle est continue en tout point $x_0 \in A$.

Autrement dit, une fonction est continue en un point x_0 si et seulement si la limite pour $x \to x_0$ de $f(x)$ existe et l'on a $\lim_{x\to x_0} f(x) = f(x_0)$.

La continuité d'une fonction est immédiate à établir lorsque celle-ci est lipschitzienne :

Définition 3.12. *On dit qu'une fonction entre deux espaces métrique $f: (X, d_X) \to (Y, d_Y)$ est lipschitzienne s'il existe $k > 0$ telle que, pour tout $x, x' \in X$ on a*

$$d_Y(f(x), f(y)) \leq k\, d_X(x, y).$$

La définition de continuité (en tout point x_0) est satisfaite, dans ce cas, avec $\delta = \epsilon/k$.

Proposition 3.9 (Critère de continuité par les suites)**.** *Une fonction $f: X \to Y$ est continue en $x_0 \in X$ si et seulement si pour toute suite $(x_n) \subset X$ convergente vers x_0 la suite $(f(x_n)) \subset Y$ converge vers $f(x_0)$.*

On voit alors que la continuité d'une fonction n'est pas affectée si l'on remplace la distance d_X par une distance équivalente sur X (ou la distance d_Y par une autre distance équivalente sur Y).

Proposition 3.10. *Soit $f: X \to Y$ une application continue entre deux espaces métriques et U un ouvert de Y. Alors $f^{-1}(U)$ (qui par définition est l'ensemble $f^{-1}(U) = \{x \in X : f(x) \in U\}$) est un ouvert de X.*

Réciproquement, si $f: X \to Y$ est une application entre espaces métriques ayant la propriété que $f^{-1}(U)$ est ouvert de X , alors f est continue dans X.

Dém. Démonstration* □

Par passage au complémentaire on en déduit la proposition suivante:

Proposition 3.11. *Soit $f: X \to Y$ une application entre deux espaces métriques. Alors f est continue si et seulement si, quel que soit F fermé dans Y, on a que $f^{-1}(F)$ est fermé dans X.*

Proposition 3.12 (Composition de fonctions continues)**.** *Si $f: X \to Y$ et $g: Y \to Z$ sont des applications continues entre espaces métriques alors l'application composée $h = g \circ f: X \to Y$ est continue.*

Dém. Il suffit d'observer que, pour tout ouvert U de Z, $h^{-1}(U) = g^{-1}(f^{-1}(U))$ est ouvert dans X par la proposition (3.10). □

Une petite variante de la dernière proposition (avec les mêmes notations) est celle-ci (la démonstration repose sur la notion de voisinage):
Si f est continue en x_0 et g est continue en $y_0 = f(x_0)$, alors la fonction $h(x) = g(f(x))$ est continue en x_0.

---

*Soit $x_0 \in f^{-1}(U)$. Alors $f(x) \in U$ qui est ouvert et donc il existe $\epsilon > 0$ tel que $B(f(x), \epsilon) \subset U$. La fonction étant continue en x_0, soit $\delta > 0$ comme dans la définition 3.10 de continuité. On a alors $B(x, \delta) \subset f^{-1}(U)$, ceci montre que $f^{-1}(U)$ est ouvert.

Réciproquement, soit $x_0 \in X$. Pour montrer que f est continue en x prenons $\epsilon > 0$. Alors par l'hypothèse $f^{-1}(B(f(x_0), \epsilon)$ est un ouvert de X, contenant x. Mais alors il existe $\delta > 0$ tel que $B(x_0, \delta) \subset f^{-1}(B(f(x_0), \epsilon))$. La définition de continuité en x_0 est alors satisfaite, puisque si $d(x, x_0) < \delta$ alors $x \in f^{-1}(B(f(x_0), \epsilon))$, c'est à dire que $d(f(x), f(x_0)) < \epsilon$. □.

3.4 Le cas des fonctions $\mathbb{R}^n \to \mathbb{R}^m$

3.4.1 Représentations graphiques des fonctions $\mathbb{R}^2 \to \mathbb{R}$

Rappelons que le graphe d'une application $f\colon X \to Y$ est l'ensemble

$$G_f = \{(x, y) \in X \times Y : y = f(x)\}.$$

Dans le cas particulier $f\colon \mathbb{R} \to \mathbb{R}$, G_f est le tracé d'une courbe dans le plan. Pour les fonctions $f\colon \mathbb{R}^2 \to \mathbb{R}$, le graphe de f se représente par une surface dans l'espace : il s'agit de la surface formées par les points de coordonnées (x, y, z), où $z = f(x, y)$.

Une autre manière de visualiser les fonctions $f\colon \mathbb{R}^2 \to \mathbb{R}$ est de représenter quelques lignes de niveau: Pour les fonctions $f\colon \mathbb{R}^2 \to \mathbb{R}$, et $k \in \mathbb{R}$, on appelle *ligne de niveau* k l'ensemble $\{(x, y) \in \mathbb{R}^2 \colon f(x, y) = k\}$. Selon les valeurs de k cet ensemble pourra être vide, mais représente souvent une courbe dans le plan. Comme sur une carte topographique, en traçant un nombre suffisant de lignes de niveau on parvient à se représenter la fonctions sans faire appel à des graphiques en 3D. Si $n \geq 3$ et $f\colon \mathbb{R}^n \to \mathbb{R}$, la généralisation de cette notion est celui de *surface de niveau* : $\{x \in \mathbb{R}^n \colon f(x) = k\}$.

Remarque 3.7. Soit $f\colon \mathbb{R}^n \to \mathbb{R}$ une fonction continue et $a \in \mathbb{R}$. Alors les ensembles $\{x \in \mathbb{R}^n \colon f(x) > a\}$ et $\{x \in \mathbb{R}^n \colon f(x) < a\}$ sont ouverts de $\mathbb{R}^n$. D'autre part, les lignes de niveau et les ensembles de la forme $\{x \in \mathbb{R}^n \colon f(x) \geq a\}$ ou $\{x \in \mathbb{R}^n \colon f(x) \leq a\}$ sont des fermées. Ces affirmations sont une conséquence des propositions 3.10 et 3.11. En effet, par exemple, on peut écrire $\{x \in \mathbb{R}^n \colon f(x) > a\} = f^{-1}(]a, +\infty[)$ et remarquer que l'intervalle $]a, +\infty[$ est un ouvert de $\mathbb{R}$.

Remarque 3.8. Une fonction $f\colon \mathbb{R}^n \to \mathbb{R}^m$ est continue en x_0 si et seulement si

$$\|f(x) - f(x_0)\| \to 0 \text{ lorsque } \|x - x_0\| \to 0.$$

3.4.2 Fonction vectorielles et application partielles

Dans cette sous-section on s'intéresse à des applications $f\colon D \subset \mathbb{R}^n \to \mathbb{R}^m$. Une telle application est dite *vectorielle*, si $m \geq 2$ (et scalaire si $m = 1$). À une telle application vectorielle on lui associe m applications scalaires $f_i\colon D \to \mathbb{R}^m$, $i = 1, \dots m$ et l'on note $f = (f_1, \dots, f_m)$.

Proposition 3.13 (Continuité des fonctions à valeur dans $\mathbb{R}^m$)**.** *Soit $f = (f_1, \dots f_m)$ une application vectorielle comme ci-dessus. Alors f est continue en $x_0 \in D$ si et seulement si les m applications scalaires $f_1, \dots f_m$ sont continues.*

D'autre part, à partir d'une application (vectorielle ou scalaire) définie sur $\mathbb{R}^n$ on peut fixer des variables et obtenir ainsi d'autres applications. Considérons, par exemple, le cas d'une application $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto f(x, y)$. On lui associe les *applications partielles* $f(\cdot, y)\colon \mathbb{R} \to \mathbb{R}$, définies par $x \mapsto f(x, y)$ ainsi que les applications partielles $f(x, \cdot)\colon \mathbb{R} \to \mathbb{R}$, définies par $y \mapsto f(x, y)$. La continuité des applications partielles $f(\cdot, y_0)$ et $f(x_0, \cdot)$ ne garantit pas la continuité de l'application f en (x_0, y_0). Par exemple, l'application $f\colon \mathbb{R}^2 \to \mathbb{R}$ définie par $f(0, 0) = 0$ et $f(x, y) = xy/(x^2 + y^2)$ n'est pas continue en $(0, 0)$ bien que les applications partielles le soient.

Proposition 3.14 (Opérations avec les fonction continues à valeur réelles)**.** *Si $f, g\colon \mathbb{R}^n \to \mathbb{R}$ sont deux fonctions continues en $x_0 \in \mathbb{R}^n$, alors les fonctions $f + g$, fg sont continues en x_0. Il en est de même pour $1/f$, à condition que $f(x_0) \neq 0$.*

3.4.3 Utilisation des coordonnées polaires

Dans le cas des fonctions $f\colon \mathbb{R}^2 \to \mathbb{R}$, pour étudier la continuité en $(0,0)$ il est parfois utile de faire appel aux coordonnées polaires : posons, pour $\rho \geq 0$ et $\theta \in [0, 2\pi]$,

$$x = \rho\cos(\theta), \quad y = \rho\sin(\theta), \qquad \text{ainsi} \qquad \rho = \sqrt{x^2+y^2}.$$

À toute application $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto f(x,y)$ on lui associe une application $\tilde{f}\colon \mathbb{R}^+ \times [0, 2\pi[\to \mathbb{R}$ définie par

$$\tilde{f}(\rho,\theta) = f(\rho\cos\theta, \rho\sin\theta).$$

Les deux applications f et $\tilde{f}$ sont essentiellement la même fonction écrite avec deux systèmes de coordonnées différentes. On a

$$\begin{aligned}
\lim_{(x,y)\to(0,0)} f(x,y) = \ell &\iff \forall\,\epsilon>0\ \exists\,\rho_0>0 \text{ t.q. } (0 \leq \rho < \rho_0,\ 0 \leq \theta < 2\pi) \Rightarrow |f(\rho,\theta)-\ell| < \epsilon \\
&\iff \Big(\sup_{\theta\in[0,2\pi[} |\tilde{f}(\rho,\theta)-\ell|\Big) \to 0 \\
&\iff \lim_{\rho\to 0} \tilde{f}(\rho,\theta) = \ell \text{ uniformément en } \theta
\end{aligned}$$

Dans la pratique, on utilise l'encadré suivant:

$$\lim_{(x,y)\to(0,0)} f(x,y) = \ell \iff \begin{array}{l} \exists\, G\colon \mathbb{R}^+ \to \mathbb{R} \text{ t.q. } G(\rho)\to 0 \text{ pour } \rho \to 0 \text{ et} \\ \forall\,\theta \in [0,2\pi[, \text{ on a } \quad |\tilde{f}(\rho,\theta)-\ell| \leq G(\rho). \end{array}$$

Exemple 3.9. Démontrons que $\lim_{(x,y)\to(0,0)} x^3/(x^2+y^2) = 0$. En passant aux coordonnées polaires, après simplification du numérateur et dénominateur par ρ^2, on obtient la fonction $\tilde{f}(\rho,\theta) = \rho\cos^3\theta$. Mais $|\tilde{f}(\rho,\theta) - 0| \leq \rho \to 0$ et la remarque dans l'encadré s'applique avec $G(\rho) = \rho$ (il est important que la fonction G puisse être prise indépendante de θ).

On aurait pu trouver le même résultat, sans utiliser les coordonnées polaires, via le théorème des gendarmes : comme $0 \leq x^2/(x^2+y^2) \leq 1$, on voit que $0 \leq |f(x,y)| \leq |x|$ et cet encadrement donne immédiatement que $f(x,y) \to 0$ pour $(x,y) \to 0$.

Exemple 3.10. Démontrons que la limite $\lim_{(x,y)\to(0,0)} (xy)/(x^2+y^2) = 0$ n'existe pas. Il suffit de trouver deux suites convergentes vers l'origine, (x_n, y_n) et (x'_n, y'_n), telles que $f(x_n,y_n)$ et $f(x'_n,y'_n)$ n'ont pas la même limite. Par exemple $f(\frac{1}{n},\frac{1}{n}) = \frac{1}{2}$ et $f(\frac{1}{n},0) = 0$.

On pouvez parvenir au même résultat à l'aide des coordonnées polaires : dans ce cas $\tilde{f}(\rho,\theta) = \cos\theta\sin\theta$. Donc $\lim_{\rho\to 0} \tilde{f}(\rho,\theta)$ dépend de θ ce qui suffit pour conclure que la limite n'existe pas. Observons que la fonction f est constante le long les droites passant par l'origine.

3.4.4 Limites à l'infini des fonctions définies dans $\mathbb{R}^n$

Soit $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$, où D contient l'extérieur d'une boule. On écrit :

$$\lim_{\|x\|\to\infty} f(x,y) = \ell \iff \forall\,\epsilon > 0, \quad \exists\, R > 0 \text{ tel que } \forall\, x \text{ t.q.} \|x\| > R \text{ on a } |f(x)-\ell| < \epsilon.$$

Dans $\mathbb{R}^2$, on utilise souvent les coordonnées polaires (et dans $\mathbb{R}^3$ lescoordonnées sphériques) pour calculer ce type de limites.

Pour les fonctions à valeur dans $\mathbb{R}$, on peut facilement donner un sens aux écritures de type $\lim_{x\to x_0} f(x,y) = +\infty$, ou encore $\lim_{\|x\|\to\infty} f(x,y) = +\infty$, etc.

3.5 Bornes supérieures, inférieures

Rappelons que si $A \subset \mathbb{R}$ est une partie bornée supérieurement, alors la borne supérieure de A est le plus petit majorant de A :

$$S = \sup A \iff \begin{array}{l} \forall a \in A,\ a \leq S \text{ et} \\ \exists (a_n) \subset A \text{ telle que } a_n \to S. \end{array}$$

On dit que le $\sup(A)$ *est atteint* lorsque $S \in A$. On dit alors que $\sup(A)$ est *le maximum* de A et on écrit plutôt $S = \max A$. Si A n'est pas borné supérieurement on pose $\sup A = +\infty$. Si $A \subset \mathbb{R}$ est une partie bornée inférieurement, alors la borne inférieure de A est le plus grand minorant de A :

$$I = \inf A \iff \begin{array}{l} \forall a \in A,\ I \leq a \text{ et} \\ \exists (a_n) \subset A \text{ telle que } a_n \to I. \end{array}$$

On dit que l'$\inf(A)$ *est atteint* lorsque $I \in A$. On dit alors que $\inf(A)$ est *le minimum* de A et on écrit $I = \min A$. Si A n'est pas borné inférieurement, on pose $\inf A = -\infty$. Une propriété profonde de $\mathbb{R}$ (assez subtile à démontrer à partir de la définition rigoureuse de $\mathbb{R}$) est que *les bornes supérieure et inférieure de tout ensemble borné existent* et sont bien des nombres réels*.

Si $f\colon D \to \mathbb{R}$, nous notons indifférement $\sup f = \sup_{x\in D} f(x) = \sup f(D)$ et $\inf f = \inf_{x \in D} f(x) = \inf f(D)$.

3.6 Ensembles compacts et théorème de Weierstrass

Définition 3.13 (Compacts)**.** *Dans un espace métrique on dit qu'un ensemble K d'un espace métrique X est* (séquentiellement) compact *si et seulement si de toute suite (x_n) contenue dans K on peut extraire une sous-suite convergente, avec limite dans K.*

Exemple 3.11.

- Dans $\mathbb{R}$, tout intervalle $[a, b]$ est compact. (Ce n'est pas immédiat à voir. Cette affirmation est connue sous le nom de "théorème de Bolzano–Weierstrass").
- L'ensemble $]0, 1]$ n'est pas compact : en effet, la suite $(1/n)_{n\in\mathbb{N}^*}$ est bien contenue dans $]0, 1]$, mais il est impossible d'extraire une sous suite convergente dans $]0, 1]$, puisque toute suite extraite converge vers 0, qui n'est pas dans l'ensemble considéré.

Proposition 3.15. *Dans un espace métrique (X, d), si $K \subset X$ est compact, alors K est une partie bornée et fermée dans X.*

Dém. En effet, si par contradiction K n'était pas fermée, alors on pourrait trouver une suite $(x_n) \subset K$ telle que $d(x_0, x_n) \to +\infty$. Toute suite extraite de (x_n) serait alors non bornée, et donc non convergente. Cela contredit la compacité de K.

De plus, si K n'était pas fermée alors on pourrait trouver $x \in \overline{K}\backslash K$, et donc un suite $(x_n) \subset K$ telle que $x_n \to x$ (d'après la Proposition 3.6). Mais alors toute suite extraite de (x_n) serait convergente vers $x \notin K$. Cela crentredit encore une fois la compacité de K. □

Dans un espace métrique général, on peut trouver des parties fermées et bornées qui ne sont pas compactes. C'est typiquement le cas dans des espaces vectoriels normés de dimension infinie. La situation est différente dans $\mathbb{R}^n$, comme le théorème fondamental suivant l'affirme.

*En revanche, ça peut arriver que la borne supérieure ou inférieure d'un ensemble borné de nombre rationnels ne soit pas un nombre rationnel. Par exemple, considérer l'ensemble de nombre rationnels $A = \{\sum_{k=0}^{n} \frac{1}{k!} : n \in \mathbb{N}\}$. Pour cet ensemble on a $\inf(A) = \min(A) = 1$ et $\sup(A) = e$, qui est irrationnel (observer que la somme de la série de terme générale $\frac{1}{k!}$ est égale à e, ce qu'on démontre en appliquant la formule de Taylor à la fonction exponentielle). Pour cet ensemble, $\max(A)$ n'existe pas.

Théorème 3.16 (Borel-Lebesgue.). *Dans $\mathbb{R}^n$, un ensemble est compact si et seulement s'il est fermé et borné.*

Dém. Démonstration hors programme. $\square$

Théorème 3.17. *Si $f\colon X \to Y$ est une fonction continue entre deux espaces métriques et $K \subset X$. Si K est compact, alors $f(K) = \{f(x) \in Y : x \in K\}$ est compact.*

Dém. Soit $y_n \subset f(K)$. On a $y_n = f(x_n)$, où $(x_n) \subset K$. Alors il existe une suite extraite (x_{n_k}) et $x \in K$ tels que $x_{n_k} \to x$. Mais alors $f(x_{n_k}) \to f(x) \in f(K)$. On a alors pu extraire de (y_n) une suite convergente dans $f(K)$. $\square$

Théorème 3.18 (de Weierstrass). *Soit $f\colon K \to \mathbb{R}$ une fonction continue, où K est un compact (de $\mathbb{R}^n$ ou plus en général d'un espace métrique X). Alors f est bornée et possède un minimum et un maximum absolu sur K.*

Dém. On sait déjà que $f(K)$ est compact (et donc borné). La fonction f est alors bornée sur K. Soit $S = \sup_{x\in K} f(x)$. Cela signifie que $\forall x \in K$ on a $f(x) \leq S$ et qu'il existe $(x_n) \subset K$ telle que $f(x_n) \to S$ (on appelle (x_n) une suite maximisante). Mais comme K est compact, il existe $x \in K$ et (x_{n_k}) extraite de (x_n) telle que $x_{n_k} \to x$. Alors $f(x_{n_k}) \to f(x)$ par la continuité de f. En conclusion, $S = \lim_{k\to\infty} f(x_{n_k}) = f(x)$. Mais alors $f(x) = \max_{x\in K} f(x)$. L'existence du minimum absolu se démontre de la même manière. $\square$

4 Dérivées partielles et différentiabilité des fonctions de plusieurs variables

4.1 Dérivées partielles

Soit $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ et $x_0 \in \overset{\circ}{D}$. Une *direction dans* $\mathbb{R}^n$ est un vecteur $v \in \mathbb{R}^n$ tel que $\|v\| = 1$. La droite d passant par x_0 et de direction v est l'ensemble des point $d = \{x_0 + tv \in \mathbb{R}^n : t \in \mathbb{R}\}$. Si l'on restreint la fonction f à cette droite on obtient une fonction d'une seule variable (la variable $t \in \mathbb{R}$): $\phi(t) = f(x_0 + tv)$. Le calcul de $\phi'(t)$ nous renseigne sur le taux d'accroissement de f le long de cette droite. Par définition, la *dérivée de f le long de la direction v en x_0* est $\phi'(0)$ (à condition que cette dérivée existe). Autrement dit:

$$\frac{\partial f}{\partial v}(x_0) := \lim_{t\to 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

Si la limite ci-dessus existe on dira que f est dérivable en x_0 le long de la direction v.

Exemple 4.1. Considérons dans $\mathbb{R}^3$ la fonction $f(x, y, z) = xyz$. Soit v la direction donnée par le vecteur de norme unitaire $v = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$. Au point $(1, 2, 3) \in \mathbb{R}^3$, on a

$$\frac{\partial f}{\partial v}(1, 2, 3) = \lim_{t\to 0} \frac{(1 + \frac{t}{\sqrt{2}})(2 + \frac{t}{\sqrt{2}})3 - 6}{t} = \frac{6}{\sqrt{2}}.$$

Dans $\mathbb{R}^n$, on note généralement par $x_1, \ldots x_n$ les n variables et $e_1 = (1, 0, \ldots, 0)$, $\ldots$, $e_n = (0, \ldots, 0, 1)$ les directions correspondant aux n-axes. Les dérivées directionnelles le long de ces vecteurs s'appellent *dérivées partielles* et on utilise les notations

$$\frac{\partial f}{\partial x_i}(x_0) = f_{x_i}(x_0) = D_i(f)(x_0) = \frac{\partial f}{\partial e_i}(x_0), \qquad i = 1, \ldots, n.$$

Dans $\mathbb{R}^2$ et $\mathbb{R}^3$ on préfère souvent appeler x, y et respectivement x, y, z les variables. On utilise alors, par exemple, la notation

$$\frac{\partial f}{\partial x}(x_0, y_0) = f_x(x_0, y_0) = D_x(f)(x_0, y_0)$$

pour la première dérivée partielle d'une fonction de deux variables au point $(x_0, y_0) \in \mathbb{R}^2$.

Exemple 4.2. Calculons les dérivées partielles $f_x(x, y)$ et $f_y(x, y)$ en tout point $(x, y) \in \mathbb{R}^2$, pour la fonction $f(x, y) = x \sin(xy)$:

$$f_x(x, y) = \sin(xy) + xy \cos(xy), \qquad \text{et} \qquad f_y(x, y) = x^2 \cos(xy).$$

Remarque 4.3. L'existence des dérivées partielles d'une fonction en un point ne garantit pas la continuité de la fonction en ce point. Considerons par exemple la fonction $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ définie par

$$f(x, y) = \left(\frac{x^2 y}{x^4 + y^2}\right)^2, \quad (x, y) \neq (0, 0), \qquad f(0, 0) = 0.$$

On peut vérifier que cette fonction possède les dérivées partielles en $(0, 0)$, et même des dérivées le long toutes les direction à l'origine. Cependant cette fonction n'est pas continue en $(0, 0)$ (comme on le voit en étudiant la suite $f(1/n, 1/n^2)$). Cela est en contraste avec les fonctions réelles d'une seule variable, pour lequelles la dérivabilité implique la continuité.

Ce n'est donc pas l'existence des dérivées partielle qui implique la continuité d'une fonction, mais la notion plus restrictive de différentiabilité, définie ci-après.

4.2 Fonctions différentiables

Le cas des fonctions d'une seule variables. Rappels. Rappelons qu'une fonction $f \colon \mathbb{R} \to \mathbb{R}$ est dérivable en un point $x_0 \in \mathbb{R}$ si et seulement s'il existe un nombre réel, noté $f'(x_0)$ (et appélé dérivée de f en x_0) tel que

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = 0 \qquad (\text{ici } h \in \mathbb{R}).$$

Pour les fonctions $f \colon \mathbb{R} \to \mathbb{R}$ dérivables en un point x_0, on peut alors donner un sens à la notion de *droite tangente* au graphe de la fonction f en x_0: il s'agit de la droite

$$d = \{(x, y) \in \mathbb{R}^2 \colon\ y = f(x_0) + f'(x_0)(x - x_0)\}.$$

Généralisons cette idée aux fonctions de plusieurs variables.

Le cas des fonctions de plusieurs variables. Rappelons qu'une *application linéaire* $L \colon \mathbb{R}^n \to \mathbb{R}$ est une application de la forme

$$L(x) = a_1 x_1 + \ldots + a_n x_n, \qquad \text{pour tout } x = (x_1, \ldots x_n) \in \mathbb{R}^n,$$

où $a_1, \ldots a_n$ sont des nombres réels. Le graphe d'une telle application est une droite de $\mathbb{R}^2$ (si $n = 1$) ou un plan de $\mathbb{R}^3$ (si $n = 2$) passant par l'origine. Plus en général, le graphe de L est un *hyperplan* de $\mathbb{R}^{n+1}$ passant par l'origine: $\{x \in \mathbb{R}^{n+1} \colon x_{n+1} = L(x_1, \ldots x_n) \colon x \in \mathbb{R}^n\}$.

Une *fonction affine* : $\mathbb{R}^n \to \mathbb{R}$ est une fonction de la forme

$$x \mapsto \alpha + L(x), \qquad x \in \mathbb{R}^n,$$

où L est une application linéaire. L'addition du terme $\alpha \in \mathbb{R}$ ayant l'effet d'une translation, le graphe des applications affines sont des hyperplans de $\mathbb{R}^{n+1}$ (ne passant pas par l'origine en général, mais plutôt par le point $(0, \ldots, 0, \alpha)$).

Définition 4.1. *Soit $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ et $x_0 \in \overset{\circ}{D}$. On dit que F est différentiable au point x_0 s'il existe une application linéaire L (dépendente de x_0) telle que*

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)-L(h)}{\|h\|} = 0 \qquad \text{(ici } h \in \mathbb{R}^n\text{)}.$$

Si f est différentiable en tout point $x_0 \in D$ on dit que f est différentiable en D. L'application L s'appelle différentielle de f en x_0. *Cette application linéaire L est notée $\mathrm{d}f_{x_0}$.*

Il est utile de comparer les deux notions suivantes :

f continue en $x_0 \iff f(x_0+h) = f(x_0) + o(1)$ pour $\|h\| \to 0$.
f différentiable en $x_0 \iff f(x_0+h) = f(x_0) + \mathrm{d}f_{x_0}(h) + o(\|h\|)$ pour $\|h\| \to 0$.

Ici $o(1)$ désigne une fonction (de la variable $h \in \mathbb{R}^n$) de limite nulle pour $\|h\| \to 0$, et $o(\|h\|)$ désigne une fonction $\varepsilon(h)$ telle que $\lim_{h\to 0} \frac{\varepsilon(h)}{\|h\|} = 0$.

Proposition 4.1. *Soit f une fonction différentiable en x_0. Alors f possède de dérivées le long toute direction et*

$$\frac{\partial f}{\partial v}(x_0) = \mathrm{d}f_{x_0}(v) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(x_0). \tag{4.1}$$

Dém. Soit $v \in \mathbb{R}^n$, une direction, On a, grâce à la linéarité de $\mathrm{d}f_{x_0}$, $\mathrm{d}f_{x_0}(tv) = t\,\mathrm{d}f_{x_0}(v)$. Ainsi,

$$\frac{f(x_0+tv)-f(x_0)}{t} = \frac{\mathrm{d}f_{x_0}(tv) + o(\|tv\|)}{t} = \mathrm{d}f_{x_0}(v) + o(1), \qquad \text{pour } t \to 0.$$

Mais alors, en prenant $t \to 0$ dans cette expression, on trouve $\frac{\partial f}{\partial v}(x_0) = \mathrm{d}f_{x_0}(v)$. En développant v avec la base caninique, $v = \sum_{i=1}^n v_i e_i$, la dernière égalité de (4.1) en découle. □

Si f est différentiable en x_0 la différentielle de f en x_0 est unique, comme on le voit grâce à l'équation (4.1).

Pour une fonction différentiable en x_0, le graphe de la fonction : $\mathbb{R}^n \to \mathbb{R}$, définie par $h \mapsto f(x_0) + \mathrm{d}f_{x_0}(h)$ étant un hyperplan, la notion de différentiabilité signifie précisément qu'au voisinage de $(x_0, f(x_0)) \in \mathbb{R}^{n+1}$, le graphe de la fonction f est convenablement approché par un hyperplan. Si l'on observe que

f différentiable en $x_0 \iff f(x) = f(x_0) + \mathrm{d}f_{x_0}(x - x_0) + o(\|x - x_0\|)$ pour $x \to x_0$,

cela conduit à la définition suivante.

Définition 4.2. *Pour une fonction différentiable en $x^0 \in \overset{\circ}{D}$, l'hyperplan tangent au graphe de la fonction $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ est l'hyperplan d'équation*

$$\Big\{x \in \mathbb{R}^{n+1} \colon \; x_{n+1} = f(x^0) + \sum_{i=1}^n (x_i - x_i^0) \frac{\partial f}{\partial x_i}(x_0)\Big\}.$$

Exemple 4.4. Construisons, dans $\mathbb{R}^3$, le plan tangent au graphe de la fonction de deux variables $f(x,y) = x^2 + \sin y + 10$ au point $(1,0)$. On a $f_x(1,0) = 2$ et $f_y(1,0) = 1$. De plus, $f(1,0) = 11$. Le plan cherché est donc le plan d'équation $z = 2(x-1) + y + 11$, ou encore $z = 2x + y + 9$.

Exemple 4.5. La différentielle $\mathrm{d}L_{x_0}$ d'une application linéaire $L\colon \mathbb{R}^n \to \mathbb{R}$ est indépendante de x_0 et coïncide avec l'application L. En particulier, la différentielle de l'application linéaire (de "projection sur la i-ème composante") $x \mapsto x_i$, que l'on note abusivement x_i, coïncide avec l'application elle même). Cela explique pourquoi on note $\mathrm{d}x_i$ l'application

$$\mathrm{d}x_i \colon \mathbb{R}^n \to \mathbb{R}, \qquad \text{où } \forall h \in \mathbb{R}^n,\ \mathrm{d}x_i(h) = h_i.$$

Pour une fonction f, la différentielle de f, notée $\mathrm{d}f$ est l'application $x_0 \mapsto \mathrm{d}f_{x_0}$. C'est pourquoi l'on peut écrire l'égalité d'applications linéaires

$$\forall\, x_0 \quad \mathrm{d}f(x_0) = \mathrm{d}f_{x_0} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)\mathrm{d}x_i,$$

(si on applique cette formule à $v \in \mathbb{R}^n$ on trouve exactement la formule (4.1)) ou en abrégé

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\mathrm{d}x_i.$$

Définition 4.3 (Gradient)**.** *Pour une fonction $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$, admettant toutes les dérivées partielles en $x_0 \in \overset{\circ}{D}$, le* gradient *de f en x_0, noté $\nabla f(x_0)$ est le vecteur de $\mathbb{R}^n$ de composantes*

$$\nabla f(x_0) = \Big(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0)\Big)^T.$$

Le gradient est par convention un vecteur colonne, c'est pourquoi on prend ici la transposition du vecteur $\big(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0)\big)$.

Observons que la relation (4.1) s'écrit

$$\mathrm{d}f_{x_0}(v) = \frac{\partial f}{\partial v}(x_0) = \langle \nabla f(x_0), v\rangle. \tag{4.2}$$

Autrement dit, la dérivée le long d'une direction v s'obtient en prenant le produit scalaire entre le vecteur gradient et le vecteur v.

Pour une fonction f et un point x_0 donné, calculons $\frac{\partial f}{\partial v}(x_0)$ le long différentes directions v: d'après (4.2) et l'inégalité de Cauchy-Schwarz (et le fait que $\|v\| = 1$),

$$-\|\nabla f(x_0)\| \leq \frac{\partial f}{\partial v}(x_0) \leq \|\nabla f(x_0)\|.$$

Les égalité sont possibles lorsque v est parallèle au vecteur $\nabla f(x_0)$. Plus précisément, si le vecteur gradient ne s'annule pas en x_0,

$$\begin{aligned} v =& \nabla f(x_0)/\|\nabla f(x_0)\| \quad &\Rightarrow \quad \frac{\partial f}{\partial v}(x_0) = \|\nabla f(x_0)\|,\\ v =& -\nabla f(x_0)/\|\nabla f(x_0)\| \quad &\Rightarrow \quad \frac{\partial f}{\partial v}(x_0) = -\|\nabla f(x_0)\|. \end{aligned}$$

Mais $\frac{\partial f}{\partial v}(x_0)$ exprime la pente du graphe de f le long de la direction v. Donc, au point x_0 la pente de la surface qui represente la fonction f sera maximale (et positive) dans la direction du gradient, minimale (et négative) si on prend l'orientation opposée à $\nabla f(x_0)$.

Par la formule (4.2) on définit

$$\frac{\partial f}{\partial w}(x_0) = \langle \nabla f(x_0), w\rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)w_i$$

pour tout vecteur de $\mathbb{R}^n$, $w \neq 0$ (non nécessairement de norme $= 1$).

Théorème 4.2 (de la différentielle totale)**.** *Si $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ est une application admettant toutes les dérivées partielles au voisinage d'un point x^0 intérieur à D et que celles-ci sont continues en x_0, alors f est différentiable en x^0.*

Dém. (dans le cas des fonctions de deux variables.) Pour $x = (x_1, x_2)$ et $x^0 = (x_1^0, x_2^0)$, en appliquant le théorème des accroissements finis, on peut trouver un réel ξ_1 compris entre x_1^0 et x_1 et un réel ξ_2 compris entre x_2 et x_2^0 tels que

$$\begin{aligned} f(x) - f(x^0) &= f(x_1, x_2) - f(x_1^0, x_2) + f(x_1^0, x_2) - f(x_1^0, x_2^0) \\ &= D_1(f)(\xi_1, x_2)(x_1 - x_1^0) + D_2(f)(x_1^0, \xi_2)(x_2 - x_2^0). \end{aligned}$$

Mais alors

$$\begin{aligned} &f(x) - f(x^0) - D_1(f)(x^0)(x_1 - x_1^0) - D_2(f)(x^0)(x_2 - x_2^0) \\ &\qquad = [D_1(f)(\xi_1, x_2) - D_1(f)(x^0)](x_1 - x_1^0) + [D_2(f)(x_1^0, \xi_2) - D_2(f)(x^0)](x_2 - x_2^0) \\ &\qquad = o(\|x - x^0\|) \qquad \text{pour } x \to x_0, \end{aligned}$$

car $D_1(f)(\xi_1, x_2) \to D_1(f)(x^0)$ et $D_2(f)(x_1^0, \xi_2) \to D_2(f)(x^0)$ grâce à la continuité des dérivées partielles de f, ce qui implique que les deux termes dans les crochets sont de limite nulle. □

4.3 Dérivées successives

Notons, pour une fonction définie sur un ouvert de $\mathbb{R}^2$, $\frac{\partial^2 f}{\partial x \partial y} = (f_y)_x$ et $\frac{\partial^2 f}{\partial y \partial x} = (f_x)_y$.

Théorème 4.3 (Théorème de Schwarz. Démonstration hors programme)**.** *Soit U un ouvert de $\mathbb{R}^2$ et $f\colon U \to \mathbb{R}$ et $(x_0, y_0) \in U$. Si les dérivées mixtes $\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$ et $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$ existent dans U et sont continues dans en (x_0, y_0), alors elles coïncident.*

Plus en général, si f est une fonction définie sur un ouvert U de $\mathbb{R}^n$, on dit qu'elle est de classe C^k dans U si toutes les dérivées itérées de la fonction f existent et sont continues dans U, jusqu'à l'ordre k. On peut généraliser le theorème de Schwarz de la manière suivante : pour une fonction de classe C^k toutes les dérivées itérées jusqu'à l'ordre k coïncident.

4.4 La différentielle des fonctions $f\colon \mathbb{R}^n \to \mathbb{R}^m$. Matrice jacobienne

Désignons par $(e_1, \dots e_n)$ et $(e_1, \dots e_m)$ les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$. Rappelons qu'à toute application linéaire $L\colon \mathbb{R}^n \to \mathbb{R}^m$ on associe une (et une seule) matrice A, de taille, $m \times n$, par la relation

$$L(x) = Ax, \qquad \text{où} \qquad Ax = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j} x_j \\ \vdots \\ \sum_{j=1}^n a_{mj} x_j \end{pmatrix}$$

Si $L\colon \mathbb{R}^n \to \mathbb{R}^m$, où $L = (L_1, \dots L_m)$ est une application linéaire, la matrice $m \times n$ qui représente L est la matrice où le vecteur $L(e_j) \in \mathbb{R}^m$, a pour composantes $a_{ij} = L_i(e_j)$, pour $i = 1, \dots, n$.

Exercice 4.6. Démontrer que, pour toute matrice A de taille $m \times n$, il existe une constante $K \geq 0$ telle que $\|Ax\| \leq K\|x\|$. Conclure que, si $L(x) = Ax$, alors l'application linéaire $L\colon \mathbb{R}^n \to \mathbb{R}^m$ est K-lipschitzienne (et donc continue).*

*Solution. $|(Ax)_i| = |\sum_{j=1}^n a_{ij} x_j| \leq \sqrt{a_{i1}^2 + \dots + a_{i,n}^2}\|x\|$ par l'inégalité de Cauchy-Schwarz. Mais alors $\|Ax\| = \sqrt{\sum_{i=1}^m (Ax)_i^2} \leq \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}\,\|x\|$. Il suffit alors de prendre $K = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$. Maintenant, d'après la linéarité de L, $\|L(x) - L(y)\| = \|L(x - y)\| \leq K\|x - y\|$. Cela montre que L est lipschitzienne.

Définition 4.4. *Soit $f\colon D \subset \mathbb{R}^n \to \mathbb{R}^m$ et $x_0 \in \overset{\circ}{D}$. On dit que F est différentiable au point x_0 s'il existe une application linéaire, $L\colon \mathbb{R}^n \to \mathbb{R}^m$ telle que*

$$\lim_{h\to 0} \frac{f(x_0+h) - f(x_0) - L(h)}{\|h\|} = 0 \qquad \textit{(ici } h \in \mathbb{R}^n\textit{)}.$$

L'application L s'appelle différentielle de f en x_0. *Cette application linéaire L est notée $\mathrm{d}f_{x_0}$. Ainsi, $\mathrm{d}f_{x_0}\colon \mathbb{R}^n \to \mathbb{R}^m$.*

On peut aussi écrire la limite précédente par

$$f(x+h) = f(x_0) + \mathrm{d}f_{x_0}(h) + o(\|h\|), \qquad \text{pour } h \to 0. \tag{4.3}$$

Observons que $f = (f_1, \dots f_m)$ et $L = (L_1, \dots, L_m)$, les composantes de L étant des applications linéaires : $\mathbb{R}^n \to \mathbb{R}$. De plus, la limite précédente (qui est une limite dans $\mathbb{R}^m$) est nulle si et seulement si elle est nulle composante par composante. Autrement dit, f est différentiable si et seulement si ses composantes $f_j\colon \mathbb{R}^n \to \mathbb{R}$ sont des applications différentiables. On peut aussi expliciter les composantes de $\mathrm{d}f_{x_0}\colon \mathbb{R}^n \to \mathbb{R}^m$ en écrivant

$$\mathrm{d}f_{x_0} = \big(\mathrm{d}(f_1)_{x_0}, \dots, \mathrm{d}(f_m)_{x_0}\big).$$

Appliquons ceci à $\mathrm{d}f_{x_0}$. Il s'agit de calculer $\mathrm{d}(f_i)_{x_0}(e_i)$. Mais $\mathrm{d}(f_i)_{x_0}(e_j) = \frac{\partial f_i}{\partial x_j}(x_0)$ d'après (4.1). En conclusion, la matrice représentative de la différentielle d'une application $f\colon \mathbb{R}^n \to \mathbb{R}^m$ (en un point x_0 donné) est la *matrice jacobienne*, donnée par :

$$J_f(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Ainsi, la formule (4.1) se généralise aux fonctions à valeurs vectorielles par

$$\mathrm{d}f_{x_0}(v) = J_f(x_0)v, \tag{4.4}$$

le vecteur à droite dans (4.4) étant le produit entre la matrice Jacobienne de f en x_0 et le vecteur v (vu comme vecteur colonne).

Si $f\colon \mathbb{R}^n \to \mathbb{R}^m$ est différentiable en un point $x_0 \in \mathbb{R}^n$, observons que l'on a l'inégalité suivante qui découle de l'exercice 4.6:

$$\exists K \geq 0 \text{ t.q. } \forall v \in \mathbb{R}^n, \quad \|\mathrm{d}f_{x_0}(v)\| \leq K\|v\|. \tag{4.5}$$

Le fait de choisir une norme non euclidienne a pour seule conséquence de modifier la valeur de la constante M.

Théorème 4.4. *Soient $f\colon \mathbb{R}^n \to \mathbb{R}^m$ et $g\colon \mathbb{R}^m \to \mathbb{R}^k$ deux fonctions telles que f est différentiable en $x_0 \in \mathbb{R}^n$ et g est différentiable en $y_0 = f(x_0) \in \mathbb{R}^m$. Alors la fonction composée $F = g \circ f\colon \mathbb{R}^n \to \mathbb{R}^k$ est différentiable en x_0. De plus, on a l'égalité (entre applications linéaires : $\mathbb{R}^n \to \mathbb{R}^k$) :*

$$\mathrm{d}F_{x_0} = \mathrm{d}g_{y_0} \circ \mathrm{d}f_{x_0} \tag{4.6}$$

et les matrices jacobiennes correspondantes vérifient la relation

$$J_F(x_0) = J_g(y_0) \cdot J_f(x_0). \tag{4.7}$$

Démonstration* hors programme.

Exemple 4.7. Soit $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ la fonction $f(x,y) = (x^2, y^2)$ et $g\colon \mathbb{R}^2 \to \mathbb{R}^2$ la fonction $g(u,v) = (e^{uv}, uv)$. Écrivons d'abord la matrice jacobienne de la fonction composée $g \circ f$ en appliquant le théorème 4.4. Retrouvons ensuite le même résultat en explicitant la fonction composée $g \circ f\colon \mathbb{R}^2 \to \mathbb{R}^2$.

On a

$$J_f(x,y) = \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix}, \qquad J_g(u,v) = \begin{pmatrix} ve^{uv} & ue^{uv} \\ v & u \end{pmatrix}$$

Donc,

$$J_{g\circ f}(x,y) = J_g(x^2,y^2)J_f(x,y) = \begin{pmatrix} y^2e^{x^2y^2} & x^2e^{x^2y^2} \\ y^2 & x^2 \end{pmatrix} \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix} = \begin{pmatrix} 2xy^2e^{x^2y^2} & x^2e^{x^2y^2} \\ y^2 & 2x^2y \end{pmatrix}.$$

La seconde méthode est plus rapide : on a $(g \circ f)(x,y) = (e^{x^2y^2}, x^2y^2)$, comme on le voit en imposant $u = x^2$ et $v = y^2$. On retrouve alors facilement le résultat précédent en calculat les deux dérivées partielles de la dernière fonction vectorielle.

Cas particuliers.

i) L'équation (4.7) permet de retrouver la formule classique de la dérivée de la fonction composée pour les fonctions d'une seule variable.

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{f} & \mathbb{R} & \xrightarrow{g} & \mathbb{R} \\ t & \mapsto & f(t) & \mapsto & g(f(t)). \end{array}$$

*On a

$$\begin{aligned} f(x_0+h) &= f(x_0) + \mathrm{d}f_{x_0}(h) + o(\|h\|), && \text{pour } h \to 0, \\ g(y_0+k) &= g(y_0) + \mathrm{d}g_{y_0}(k) + o(\|k\|), && \text{pour } k \to 0. \end{aligned}$$

Et, pour $h \to 0$,

$$\begin{aligned} F(x_0+h) = g(f(x_0+h)) &= g\big(f(x_0) + \underbrace{\mathrm{d}f_{x_0}(h) + o(\|h\|)}_{:=k(h)}\big) \\ &= g(y_0) + \mathrm{d}g_{y_0}(k(h)) + o(\|k(h)\|) \\ &= g(y_0) + \mathrm{d}g_{y_0}\big(\mathrm{d}f_{x_0}(h)\big) + \mathrm{d}g_{y_0}\big(o(\|h\|)\big) + o(\|k(h)\|), && \text{(par linéarité de } \mathrm{d}g_{y_0}) \\ &= g(y_0) + \mathrm{d}g_{y_0}\big(\mathrm{d}f_{x_0}(h)\big) + o(\|h\|). \end{aligned} \tag{4.8}$$

Détaillons la dernière égalité : il s'agit de démontrer que $\mathrm{d}g_{y_0}\big(o(h)\big) + o(k(h)) = o(h)$ pour $h \to 0$, c'est à dire,

$$\lim_{h\to 0} \frac{\mathrm{d}g_{y_0}\big(o(h)\big) + o(\|k(h)\|)}{\|h\|} = 0.$$

En effet, en appliquant l'inégalité (4.5) (avec g à la place de f, y_0 à la place de x_0 et M' à la place de M) on trouve, pour $h \to 0$,

$$\frac{\|\mathrm{d}g_{y_0}\big(o(h)\big)\|}{\|h\|} \le M' \frac{\|o(\|h\|)\|}{\|h\|} \to 0.$$

De plus,

$$\frac{o(\|k(h)\|)}{\|h\|} = \frac{o(\|k(h)\|)}{\|k(h)\|} \cdot \frac{\mathrm{d}f_{x_0}(h) + o(\|h\|)}{\|h\|} \to 0,$$

grâce à (4.5), parce que le premier facteur converge vers 0. Cela prouve la dernière égalité dans (4.8).

Maintenant, (4.8) signifie précisément que F est différentiable en x_0, avec différentielle donnée par la formule (4.6). La formule (4.7), découle alors de la propriété bien connue des applications linéaires, que la matrice représentative d'une application linéaire $L_1 \circ L_2$ (composé de deux application linéaires L_1 et L_2 est donné par le produit entre la matrice de l'application L_1 avec la matrice de l'application L_2. □

Dans ce cas, la composée est $F\colon \mathbb{R} \to \mathbb{R}$ et $F(t) = g(f(t))$. De plus, pour une fonction $: \mathbb{R} \to \mathbb{R}$ la matrice jacobienne (1×1) est simplement la dérivée de la fonction. Ainsi, l'égalité matricielle (4.7) n'est rien d'autre que l'égalité classique

$$F'(t) = g'(f(t))f'(t).$$

ii) Considérons un cas un peu plus compliqué :

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{f=(f_1,f_2)} & \mathbb{R}^2 & \xrightarrow{g} & \mathbb{R} \\ t & \mapsto & (f_1(t), f_2(t)) & \mapsto & g(f_1(t), f_2(t)). \end{array}$$

Dans ce cas, la composée est $F\colon \mathbb{R} \to \mathbb{R}$, où $F(t) = g(f_1(t), f_2(t))$. Observons qu'ici g est une fonctions de deux variables. Appelons (x, y) les variables de g. L'égalité matricielle (4.7) équivaut à l'égalité

$$F'(t) = \frac{\partial g}{\partial x}(f_1(t), f_2(t))f_1'(t) + \frac{\partial g}{\partial y}(f_1(t), f_2(t))f_2'(t).$$

Exemple 4.8. Soit $g = g(x, y)$ une fonction différentiable $: \mathbb{R}^2 \to \mathbb{R}$. Considérons la composition de g avec la fonction $t \mapsto (\sin t, t^3)$ et calculons ensuite la dérivée :

$$\frac{\mathrm{d}}{\mathrm{dt}}\big[g(\sin t, t^3)\big] = \frac{\partial g}{\partial x}(\sin t, \ln t)\cos t + 3\frac{\partial g}{\partial y}(\sin t, \ln t)t^2.$$

iii) Considérons un autre cas :

$$\begin{array}{ccccc} \mathbb{R}^2 & \xrightarrow{f=(f_1,f_2)} & \mathbb{R}^2 & \xrightarrow{g} & \mathbb{R} \\ (x,y) & \mapsto & (f_1(x,y), f_2(x,y)) & \mapsto & g(f_1(x,y), f_2(x,y)). \end{array}$$

Dans ce cas, la composée est $F\colon \mathbb{R}^2 \to \mathbb{R}$, où $F(x, y) = g(f_1(x, y), f_2(x, y))$. La fonction g est une fonction de deux variables. Appelons (u, v) les variables de g pour les distinguer des variables (x, y) de la fonction f. L'égalité matricielle (4.7) équivaut au système

$$\begin{cases} \dfrac{\partial F}{\partial x}(x,y) = \dfrac{\partial g}{\partial u}(f_1(x,y), f_2(x,y))\dfrac{\partial f_1}{\partial x}(x,y) + \dfrac{\partial g}{\partial v}(f_1(x,y), f_2(x,y))\dfrac{\partial f_2}{\partial x}(x,y), \\ \dfrac{\partial F}{\partial y}(x,y) = \dfrac{\partial g}{\partial u}(f_1(x,y), f_2(x,y))\dfrac{\partial f_1}{\partial y}(x,y) + \dfrac{\partial g}{\partial v}(f_1(x,y), f_2(x,y))\dfrac{\partial f_2}{\partial y}(x,y), \end{cases}$$

Exemple 4.9. Soit $g(u, v) = uv$ et $f(x, y) = (e^x, \sin(x - y))$. Il y a deux manières de calculer la dérivée partielle de $F(x, y) = (g \circ f)(x, y)$ par rapport à x.

On peut calculer directement $F(x, y) = e^x \sin(x-y)$ et l'on trouve $\frac{\partial}{\partial x}[F(x, y)] = e^x \sin(x-y) + e^x \cos(x - y)$. Mais on pouvait aussi observer que $\frac{\partial g}{\partial u} = v$ et $\frac{\partial g}{\partial v} = u$, remplacer $(u, v) = (e^x, \sin(x - y))$ et ensuite appliquer la première équation du système ci-dessus. On aboutit au même résultat.

5 Formule de Taylor et développements limités

Soit f une fontion de classe C^2 dans un ouvert U de $\mathbb{R}^n$ et soit $x_0 \in U$, et $\delta > 0$ tel que $B(x_0, \delta) \subset U$. Si $v \in \mathbb{R}^n$ est une direction et $0 \le t \le \delta$, on pose:

$$F(t) = f(x_0 + tv).$$

La formule de Taylor centrée en 0 pour la fonction F d'ordre 2 s'écrit

$$F(t) = F(0) + F'(0)t + \frac{F''(0)}{2!}t^2 + o(t^2), \qquad \text{pour } t \to 0. \tag{5.1}$$

D'autre part,

$$F'(t) = \frac{\partial f}{\partial v}(x_0 + tv) = \sum_{i=1}^{n} D_i(f)(x_0 + tv)\, v_i,$$
$$F''(t) = \sum_{i=1}^{n} \frac{\partial D_i(f)}{\partial v}(x_0 + tv)v_i = \sum_{i,j=1}^{n} D_i D_j(f)(x_0 + tv)\, v_i v_j.$$

Cela donne,

$$f(x_0 + tv) = f(x_0) + \sum_{i=1}^{n} D_i(f)(x_0)tv_i + \sum_{i,j=1}^{n} \frac{D_i D_j(f)}{2!}(x_0)t^2 v_i v_j + o(t^2) \qquad \text{pour } t \to 0.$$

En posant $h = tv$ (et donc $|t| = \|h\|$) on obtient, pour $\|h\| \to 0$,

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^{n} D_i(f)(x_0)h_i + \frac{1}{2}\sum_{i,j=1}^{n} D_i D_j(f)(x_0)h_i h_j + o(\|h\|^2).$$

C'est la formule de Taylor d'ordre 2 pour les fonctions de plusieurs variables. Le terme à droite est le *développement limité d'ordre 2* de f centré en x_0. En observer la structure typique :

constante + partie linéaire en h + partie quadratique en h + reste.

Plus en général, si f est de classe C^k au voisinage de x_0, on pourra écrire la la formule de Taylor d'ordre k pour F au voisinage de 0 :

$$F(t) = F(0) + F'(0)t + \frac{F''(0)}{2!}t^2 + \cdots + \frac{F^{(k)}}{k!}t^k + o(t^k), \qquad \text{pour } t \to 0.$$

D'autre part, en généralisant les formules pour $F'(t)$ et $F''(t)$ on trouve

$$F^{(k)}(t) = \sum_{i_1,\ldots i_k=1}^{n} D_{i_1} D_{i_2} \ldots D_{i_k}(f)(x_0 + tv)\, v_{i_1} v_{i_2} \ldots v_{i_k}.$$

On obtient alors, pour $x \to x_0$,

$$\begin{aligned} f(x_0 + h) = f(x_0) &+ \sum_{i=1}^{n} D_i f(x_0)h_i + \frac{1}{2}\sum_{i,j=1}^{n} D_i D_j f(x_0)h_i h_j + \cdots \\ &+ \frac{1}{k!}\sum_{i_1,\ldots i_k=1}^{n} D_{i_1} D_{i_2} \ldots D_{i_k} f(x_0)\, h_{i_1} h_{i_2} \ldots h_{i_k} + o(\|h\|^k). \end{aligned} \tag{5.2}$$

Le terme à droite est le dévéloppement limité d'ordre k pour f centré en x_0.

On sait qu'une fonction d'une seule variable de dérivée identiquement nulle sur un intervalle ouvert est constante sur cet intervalle. Pour généraliser ceci aux fonctions de plusieurs variables, rappelons qu'un ouvert U de $\mathbb{R}^n$ est dit *non connexe* s'il est réunion de deux ouverts non vides : $U = V \cup W$, avec $V, W \neq \emptyset$. Dans le cas contraire, on dit que U est connexe.

Proposition 5.1. *Soit U un ouvert connexe et $f: U \to \mathbb{R}$ une fonction différentiable telle que toutes les dérivées partielles d'ordre 1 sont nulles dans U. Alors, f est constante sur U.*

Dém. Soit $x_0 \in U$. Considérons les ensemble $V = \{x \in U: f(x) = f(x_0)\}$ et $W = \{x \in U: f(x) \neq f(x_0)\}$. On a clairement $U = V \cup W$ (union disjointe), avec W ouvert (puisque f est continue et W est l'image réciproque par f de l'ouvert $\{f(x_0)\}^c$) Montrons que V est ouvert. Si $x_1 \in V$ et la boule centrée en x_1 de rayon $\delta > 0$ est contenue dans U, alors f est constante dans cette boule (parce que pour toute direction v, $\frac{\partial f}{\partial v} = 0$ dans cette boule et donc f reste constante le long toute direction v dans $B(x_1, \delta)$. Mais alors $B(x_1, \delta) \subset V$ et V est ouvert. De plus V est non vide puisque $x_0 \in V$. Mais U est connexe, et donc on a nécessairement $W = \emptyset$. Mais alors $f(x) \equiv f(x_0)$ dans U. □

5.1 L'inégalité des accroissements finis

L'inégalité des accroissements finis pour les fonctions d'une seule variable est bien connue : si $F: [a, b] \to \mathbb{R}$ est dérivable, alors $|F(b) - F(a)| \leq (\sup_{t \in [a,b]})|F'(t)|\,|b - a|$. Il existe un analogue pour les fonctions de plusieurs variables :

Proposition 5.2. *Soit U un ouvert de $\mathbb{R}^n$, $f: U \to \mathbb{R}$ une fonction de classe C^1 et $x, x' \in U$, tels que le* ségment *de x à x', qui par définition est l'ensemble*

$$[x, x'] = \{(1 - t)x + tx' : 0 \leq t \leq 1\}$$

est contenu dans U. Alors

$$|f(x) - f(x')| \leq \Big(\sup_{z \in [x,x']} \|\nabla f(z)\| \Big) \|x - x'\|.$$

Dém. Posons $F(t) = f((1 - t)x + tx')$. Alors $F(0) = f(x)$ et $F(1) = f(x')$. De plus F est de classe C^1 dans l'intervalle $[0, 1]$ et $F'(t) = \langle \nabla f((1 - t)x + tx'), x' - x \rangle$. Par l'inégalité des accroissements finis, appliquée à F, ensuite par l'inégalité de Cauchy-Schwarz,

$$|f(x) - f(x')| = |F(1) - F(0)| \leq \Big(\sup_{t \in [0,1]} |F'(t)| \Big) \leq \sup_{z \in [x,x']} \|\nabla f(z)\|\, \|x - x'\|.$$

□

6 Extrema de fonctions de plusieurs variables

6.1 Points stationnaires

Définition 6.1 (extrema locaux)**.** *Soit $f: D \subset \mathbb{R}^n \to \mathbb{R}$ et $x_0 \in D$. On dit que x_0 est un* point de maximum relatif (ou local) *pour f s'il existe un voisinage V de x_0 tel que $f(x) \leq f(x_0)$ pour tout $x \in V \cap D$. Dans ce cas on dit que la valeur $f(x_0)$ est un maximum relatif (ou local) de la fonction f.*

On dit que x_0 est un point de minimum relatif (ou local) *pour f s'il existe un voisinage V de x_0 tel que $f(x_0) \leq f(x)$ pour tout $x \in V \cap D$. Dans ce cas on dit que la valeur $f(x_0)$ est un maximum relatif (ou local) de la fonction f.*

Si on a la condition plus forte $f(x) \leq f(x_0)$ pour tout $x \in D$ on dit que x_0 est un (point de) maximum absolu (ou global) pour f dans D. De même, si $f(x_0) \leq f(x)$ pour tout $x \in D$ on dit que x_0 est un point de minimum absolu (ou global) pour f dans D.

Les point de minimum s'appellent aussi *minimiseurs* et les point de maximum *maximiseurs.*

Exemple 6.1.
(i) Étudier les extrema locaux et globaux de la fonction $f : \mathbb{R} \to \mathbb{R}$, où $f(x) = (x^2 - 1)^2$ sur $\mathbb{R}$.
(ii) Étudier les extrema locaux et globaux de même fonction f sur l'intervalle $I = [\frac{1}{2}, 2]$.
Les réponses suivantes s'obtiennent en traçant le graphe de la fonction f à l'aide d'un tableau de variations.
(i) f possède 3 extrema locaux en 0, 1 et -1, Plus précisément 0 est un point de maximum locale (mais ce n'est pas un point de maximum globale : le maximum globale de f n'existe pas). De plus ± 1 sont deux points de minimum globaux sur $\mathbb{R}$ et $\min_{\mathbb{R}} f = 0$.
(ii) Sur l'intervalle I, 1 est le seul point minimum global pour fOn a $\min_I f = 0$ De plus, 2 est le seul point de maximum globale pour f dans I et $\max_I f = 9$. D'autre part $\frac{1}{2}$ est un point de maximum local (non global) de f dans I.

Observons que dans l'exemple précédent, dans le cas (i), tous les extrema locaux ont été trouvés parmi les points où f' s'annule : c'est dû au fait que la fonction a été étudié sur un intervalle ouvert. Dans le cas (ii) ce n'est plus le cas.

Dans cette section on se focalise sur la recherche de point d'extrema *libres*, c'est-à dire les points de minimum ou maximum à l'*intérieur* de l'ensemble de définition de la fonction. Cela revient à supposer que la fonction es définie sur un ouvert.

Proposition 6.1 (CNPO ou CN1)**.** *Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$, où U est un ouvert de $\mathbb{R}^n$ et $x_0 \in U$ un extremum local (c'est à dire un maximum local ou un minimum local) pour f. Si f est dérivable le long d'une direction v, alors $\frac{\partial f}{\partial v}(x_0) = 0$. En particulier, si f est différentiable en x_0, toutes les dérivées partielles de f s'annulent en x_0 et $\nabla f_{x_0} = 0$.*

Dém. La fonction d'une seule variable $F(t) = f(x_0 + tv)$ est dérivable en $t = 0$, où elle possède un extremum local. Par le critère de Fermat, $F'(0) = 0$. Mais alors $0 = F'(0) = \frac{\partial f}{\partial v}(x_0)$. □

Un point x_0 où f est différentiable et $\nabla f(x_0) = 0$ s'appelle un *point stationnaire* ou *point critique* pour f.

Les points de minimum et maximum locaux d'une fonction différentiable $f : D \to \mathbb{R}$ sont alors à chercher parmi les point stationnaires, qui sont les solutions du système

$$\begin{cases} \frac{\partial f}{\partial x_1}(x) = 0 \\ \dots \\ \frac{\partial f}{\partial x_n}(x) = 0, \end{cases} \qquad x \in \overset{\circ}{D}, \tag{CN1}$$

ou éventuellement parmi les points de $D \setminus \overset{\circ}{D}$. Ces points sont situés sur la frontière de l'ensemble D.

Le système ci-dessus s'appelle en abrégé (CNPO) ("conditions nécessaires du premier ordre") ou (CN1), puisque ce sont les dérivées d'ordre 1 qui interviennent. Vérifier ces condition n'est pas une condition suffisante pour garantir que les solutions trouvées soient bien de minimiseurs ou de maximiseurs : par exemple, pour la fonction $f : \mathbb{R} \to \mathbb{R}$ telle que $f(x) = x^3$ ce système se réduit à la seule équation $3x^2 = 0$. La solution $x = 0$ ne correspont pas à un extremum pour f.

Observer que (CN1) est un système (non-linéaire !) de n équations et n inconnues. Il n'y a malheureusement pas de théorie générale pour trouver les solutions.

Exemple 6.2. Trouver les extrema locaux et globaux de la fonction $f(x, y) = \sqrt{1 - x^2 - y^2}$ sur son ensemble de définition.

Réponse : L'ensemble de définition est le disque $D = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \le 1\}$. Ce n'est pas un ouvert. Cherchons les points stationnaires intérieurs à D, c'est à dire dans $\overset{\circ}{D} = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 < 1\}$. L'unique solution du système (CN1) est le point (0,0). Observons que $f(0,0) = 1$. Mais on voit immédiatement que $f(x, y) \le 1$ pour tout $(x, y) \in D$, donc l'origine est un point de maximum global pour f. Il n'y a pas de point de minimum (local ou globale) à l'intérieur de D (sinon on aurait trouvé d'autres points stationnaires) ; cherchons alors le minimum de f sur la frontière, qui est le cercle $\partial D = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\}$. Mais on voit que la fonction f s'annule sur ∂D; d'autre part $f \ge 0$ dans D. La conclusion est que tous les points de ∂D sont des points de minimum global pour f.

Exemple 6.3. On cherche à construire un caisson de 20m^3. Le matériau pour le fond coûte 3 euros/ m^2, pour le couvercle 2 euros/m^2 et pour les côtés 1 euro/m^2. Quel est le caisson le moins cher ? Et le plus cher ?
Réponse :
Modélisation : Notons x(=longueur), y(=largeur), z(=hauteur) les mesures du casson en mètres. Le coût de construction est $C(x, y, z) = 3xy + 2xy + 2(xz + yz) = 5xy + 2xz + 2yz$. Il s'agit de résoudre les problèmes de minimisation et maximisation pour C "avec contrainte" : $\min\{C(x, y, z) \colon xyz = 20\}$ et $\max\{C(x, y, z) \colon xyz = 20\}$.
Élimination de la contrainte et recherche des points stationnaires : L'ensemble des points vérifiant $xyz = 20$ n'est pas un ouvert, c'est pourquoi la proposition (CN1) ne s'applique pas directement à la fonction C. On impose $z = 20/(xy)$ et on étudie la fonction

$$f(x, y) = C(x, y, \tfrac{20}{xy}) = 5xy + 40(\tfrac{1}{y} + \tfrac{1}{x}), \qquad x > 0, y > 0.$$

Nous pouvons appliquer la proposition (CN1) à la fonction f, qui est bien définie dans un ouvert. Le système $\nabla f(x, y) = 0$ possède une seule solution pour $x > 0$ et $y > 0$. Elle est donnée par $x = y = 2$ (et donc $z = 5$).
Synthèse : On construit donc un caisson de mesures $2 \times 2 \times 5$. Ce choix correspond-t-il au caisson de coût minimum ou maximum ? Le problème de minimisation est-il *bien posé* ? Et celui de maximisation ? [Voir l'exemple ci-après].

Dans la pratique, les fonction dont on étudie les problèmes d'optimisation sont souvent continues. Le théorème de Weierstrass est alors un outil essentiel pour garantir que des solutions existent. Mais ce théorème ne s'applique pas toujours (comme par exemple dans l'exemple précédent) Le théorème suivant est une variante utile.

Théorème 6.2 (Weierstrass - variante)**.** *Soit $f \colon D \subset \mathbb{R}^n \to \mathbb{R}$ une fonction continue. Si $\bar{x} \in D$ et si l'*ensemble de sous-niveau $K = \{x \in D \colon f(x) \le f(\bar{x})\}$ *est compact, alors le problème de minimisation* $\min_{x \in D} f(x)$ *possède une solution. Autrement dit, il existe $x^* \in D$ tel que $f(x^*) = \min_{x \in D} f(x)$.*

Dém. En effet, il existe $x^* \in K$ tel que $f(x^*) = \min_{x \in K} f(x)$, par le théorème de Weierstrass. Mais $f(x^*) \le f(\bar{x})$ puisque $\bar{x} \in K$. Si $x \in K$ on a $f(x) \ge f(x^*)$ par définition de x^*. Si $x \in D$ et $x \notin K$ on a $f(x) > f(\bar{x}) \ge f(x^*)$. En conclusion, $f(x^*) = \min_{x \in D} f(x)$. □

> Le cas typique d'application est celui d'une fonction $f \colon \mathbb{R}^n \to \mathbb{R}$ telle que $\lim_{\|x\| \to +\infty} f(x) = +\infty$. Une telle fonction a tous les ensembles de sous-niveau bornés (pourquoi ?). Si de plus f est continue, ses ensembles de sous-niveau sont fermés (pourquoi ?) et donc compacts. La fonction possède alors un minimum absolu.

Bien entendu on peut établir un théorème analogue pour l'existence d'un maximum absolu : il s'agit cette fois-ci de supposer que l'ensemble de "sur-niveau" $\{x\colon f(x) \geq f(\bar{x})\}$ est borné. Le cas typique d'application est celui d'une fonction continue telle que $\lim_{\|x\|\to+\infty} f(x) = -\infty$.

Exemple 6.4 (Retour à l'exemple 6.3)**.** Appliquons cette variante du théorème de Weierstrass avec $(\bar{x}, \bar{y}) = (1, 1)$ (ce choix est arbitraire). Observons que l'ensemble de sous-niveau $K = \{(x, y)\colon f(x, y) \leq f(1, 1)\} = 5xy + 40(\frac{1}{y} + \frac{1}{x}) \leq 85\}$ est compact (en effet, il est manifestement fermé et il est borné, puisque $x \geq 40/85$, $y \geq 40/85$ et $5xy \leq 85 \Rightarrow x \leq 17 \cdot 85/40$ et $y \leq 17 \cdot 85/40$). Mais alors le problème de minimisation de l'exemple 6.3 est bien posé, c'est-à dire que le caisson de coût minimum existe : c'est bien le caisson de mesures $2 \times 2 \times 5$ trouvé avant. Le problème de maximisation est mal posé : le caisson de coût maximum n'existe pas.

6.2 Fonctions quadratiques et matrice hessienne

Un peu d'algèbre linéaire. Commençons par des rappels d'algèbre linéaire. Un système linéaire homogène de n équations et n inconnues se représente par l'équation matricielle

$$Ax = 0, \qquad x = (x_1, \dots x_n)^T \in \mathbb{R}^n.$$

où A est une matrice carrée $n \times n$. Ce système possède toujours comme solutions au moins le vecteur nul $x = 0$. Il est bien connu que la condition nécessaire et suffisante pour qu'il y ait d'autres solutions $x \neq 0$ est que $\det A = 0$.

Définition 6.2. *Un nombre réel (ou complexe) λ est dit* valeur propre *d'une matrice carrée A s'il existe $w \in \mathbb{R}^n \backslash \{0\}$ tel que $Aw = \lambda w$. Le vecteur w s'appelle alors* vecteur propre *pour A.*

Observons que si λ est une valeur propre alors $(A - \lambda I)w = 0$, avec $w \in \mathbb{R}^n \backslash \{0\}$. Mais cela est possible si et seulement si

$$\det(A - \lambda I) = 0. \tag{6.1}$$

Chercher les valeurs propres revient à résoudre l'équation d'inconnue λ (6.1). En général, (6.1) est une équation polynomiale en λ, de degré n. Le terme à gauche dans (6.1) s'appelle le *polynôme caractéristique* de A. Cette équation possède alors n solutions, comptées avec leur multiplicité. Ces solutions sont éventuellement complexes.

Exemple 6.5. La matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ possède 2 valeurs propres λ_1 et λ_2. Celles-ci sont les deux les solutions d'une équation polynomiale de degré 2 :

$$0 = \det\left[\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = \det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{pmatrix} = \lambda^2 - 5\lambda - 1.$$

Donc $\lambda_{1,2} = \frac{1}{2}(5 \pm \sqrt{29})$.

Théorème 6.3 (démonstration hors programme. Voir le cours d'algèbre.)**.** *Si A est une matrice carrée* symétrique *de taille $n \times n$, c'est-à dire $a_{ij} = a_{ji}$ pour tout $i, j = 1, \dots, n$, alors les valeurs propres de A sont toutes réelles.*

Soit $A = (a_{ij})$ une matrice *symétrique* $n \times n$. Le produit entre la matrice A et le vecteur $h = (h_1, \dots h_n)^T$ est donné par le vecteur de composantes $(Ah)_i = \sum_{j=1}^n a_{ij} h_j$. Une *forme quadratique* est une application : $\mathbb{R}^n \to \mathbb{R}$ de la forme

$$h \mapsto \langle Ah, h \rangle = \sum_{i,j=1}^n a_{ij} h_i h_j.$$

Exemple 6.6. Les formes quadratiques de $\mathbb{R}^2$ sont toutes et seules les fonctions de la forme

$$(x,y) \mapsto \langle \begin{pmatrix} r & s \\ s & t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \rangle = r\,x^2 + 2s\,xy + t\,y^2.$$

Les formes quadratiques de $\mathbb{R}^3$ sont toutes et seules les fonctions de la forme

$$(x,y,z) \mapsto \langle \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rangle = a\,x^2 + 2b\,xy + 2c\,xz + d\,y^2 + 2e\,yz + f\,z^2.$$

Le cas des fonctions de plusieurs variables. Les formes quadratiques apparaissent naturellement dans la formule de Taylor d'ordre 2. En effet, soit f une fonction de classe C^2 au voisinage d'un point x_0. Considérons la *matrice hessienne de f en x_0*, c'est à dire la matrice symétrique de taille $n \times n$ des dérivées partielles secondes

$$H = H_f(x_0) = \Big(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)\Big)_{1 \le i,j \le n}.$$

Pour les fonctions $f\colon \mathbb{R} \to \mathbb{R}$, la matrice hessienne est réduite au seul nombre $f''(x_0)$.

La formule de Taylor pour f d'ordre 2 s'écrit alors

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h \rangle + \tfrac{1}{2}\langle H_f(x_0)h, h\rangle + o(\|h\|^2).$$

Si de plus x_0 est un point stationnaire, on a $\nabla f(x_0) = 0$ et donc la formule précedente devient

$$f(x_0 + h) = f(x_0) + \tfrac{1}{2}\langle H_f(x_0)h, h\rangle + o(\|h\|^2), \tag{6.2}$$

qui est la généralisation naturelle de (6.4).

Pour déterminer la nature du point stationnaire x_0 on est amené à diviser terme-à-terme par $\|h\|^2$ et prendre $\|h\| \to 0$. Cela conduit à étudier la fonction $h \mapsto \frac{\langle H_f(x_0)h,h\rangle}{\|h\|^2}$.

Plus en général, étudions alors les fonctions de type

$$F(h) = \frac{\langle Ah, h\rangle}{\|h\|^2}, \qquad h \neq 0. \tag{6.3}$$

avec $A = (a_{ij})$ matrice symétrique de taille $n \times n$.

Lemme 6.4. *Soit $F(h) = \frac{\langle Ah,h\rangle}{\|h\|^2}$ avec $A = (a_{ij})$ matrice symétrique de taille $n \times n$. Les points stationnaires de F sont des vecteurs propres de la matrice A. De plus, si v est un vecteur propre de A, alors*

$$Av = F(v)\,v,$$

c'est à dire que la valeur propre associée à v est $F(v)$.

Dém. Observons que $\frac{\partial}{\partial h_i}(\langle Ah, h\rangle) = 2\sum_{i=1}^n a_{ij}h_j$ et que $\frac{\partial}{\partial h_i}\|h\|^2 = 2h_i$. Il en découle que

$$\frac{\partial F}{\partial h_i}(h) = 2\frac{\sum_{j=1}^n a_{ij}h_j - F(h)h_i}{\|h\|^2} = 2\frac{[Ah - F(h)h]_i}{\|h\|^2}.$$

Soit v un point stationnaire de F. Alors, pour tout $i = 1, \dots, n$, on a $\frac{\partial F}{\partial h_i}(v) = 0$. Mais alors les numérateurs de l'expression précédente s'annulent et on trouve

$$Av = F(v)\,v.$$

$\square$

Lemme 6.5. *Soit $F(h) = \frac{\langle Ah,h\rangle}{\|h\|^2}$ avec $A = (a_{ij})$ matrice symétrique de taille $n \times n$. Soit λ_1 et λ_n la plus petite et la plus grande valeurs propre de la matrice A. Alors*

$$\min_{h\neq 0} F(h) = \lambda_1, \qquad et \qquad \max_{h\neq 0} F(h) = \lambda_n.$$

Dém. Observons que F est une fonction telle que, pour tout $\lambda > 0$, $F(\lambda h) = F(h)$. en particulier, en prenant $\lambda = \frac{1}{\|h\|}$ on voit que $F(h) = F(\frac{h}{\|h\|})$. Mais alors $\max_{h\neq 0} F(h) = \max\{F(\frac{h}{\|h\|})\colon h \neq 0\} = \max\{F(h)\colon \|h\| = 1\}$. De même, $\min_{h\neq 0} F(h) = \min\{F(h)\colon \|h\| = 1\}$. L'ensemble $\{h\colon \|h\| = 1\}$ est la sphère unité qui est compact dans $\mathbb{R}^n$. La fonction F étant continue sur $\{h\colon \|h\| = 1\}$, le théorème de Weierstrass implique que F possède un maximum et un minimum absolu sur cet ensemble. Par la discussion précédente, F possède un minimum et un maximum absolu sur $\mathbb{R}^n\backslash\{0\}$.

Soit donc h^* un point de minimum de F dans $\mathbb{R}^n\backslash\{0\}$. Alors h^* est un point stationnaire et par le lemme précédent $F(h^*)$ est une valeur propre de la matrice A. De plus, $Ah^* = F(h^*)h^*$.

Mais alors $\lambda_1 \leq F(h^*) = \min_{h\neq 0} F(h)$. Mais au valeur propre λ_1 correspondent un vecteur propre $v_1 \in \mathbb{R}^n\backslash\{0\}$. et l'on a $Av_1 = \lambda_1 v_1$, donc $F(v_1) = \lambda_1$. Donc $\min_{h\neq 0} F(h) \leq \lambda_1$.

Si $\hat{h}$ est un point de maximum pour F, on procède de la même manière et on trouve $F(\hat{h}) = \lambda_n$. □

6.3 Problèmes d'optimisations : conditions d'ordre 2

Le cas de fonctions d'une seule variable. Rappels. Soit $f\colon I \to \mathbb{R}$ une fonction de classe C^2 sur l'intervalle ouvert $I =]a, b[$ et x_0 un point stationnaire pour f dans $]a, b[$, c'est-à dire $f'(x_0) = 0$. Alors, par la formule de Taylor d'ordre 2,

$$f(x_0 + h) = f(x_0) + \tfrac{1}{2} f''(x_0)h^2 + o(h^2), \qquad \text{pour } h \to 0. \tag{6.4}$$

Si on divise terme-à-terme par h^2 et on prend $h \to 0$ on en déduit :

$$\tfrac{1}{h^2}(f(x_0 + h) - f(x_0)) = \tfrac{1}{2} f''(x_0) + o(1), \qquad \text{pour } h \to 0.$$

- Pour que x_0 soit un point de minimum local il est nécessaire que $f''(x_0) \geq 0$. (En effet, si x_0 est un point de minimum local, alors pour h assez petit le terme à gauche est ≥ 0).
- Pour que x_0 soit un point de minimum local il est suffisant que $f''(x_0) > 0$. (En effet, si $f''(x_0) > 0$ alors pour h assez petit $f''(x_0) + o(1) > 0$ et donc le terme à gauche est positif).
- Pour que x_0 soit un point de maximum local il est nécessaire que $f''(x_0) \leq 0$.
- Pour que x_0 soit un point de maximum local il est suffisant que $f''(x_0) < 0$.

Observer que lorsque $f''(x_0) = 0$ on ne peut rien conclure en général : on a parfois un minimum, ou un maximum, ou un point d'inflexion (considérer par exemple les cas $f(x) = x^3$, avec $x_0 = 0$). Généralisons ces considérations aux fonctions de plusieurs variables.

Le théorème suivant établit des conditions nécessaires (CN2) et des conditions suffisantes (CS2), faisant intervenir les dérivées d'ordre 2, pour qu'un point soit un minimiseur ou un maximiseur local.

Théorème 6.6 (CN2-CS2)**.** *Si $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ est une fonction de classe C^2 au voisinage de x_0 et si x_0 est un point stationnaire, alors:*

- *Pour que x_0 soit un point de minimum local il est nécessaire que tous les valeurs propres de la matrices hessiennes de f soient ≥ 0. (On dit que $H_f(x_0)$ est* semi-définie positive*).*

- *Pour que x_0 soit un point de minimum local il est suffisant que tous les valeurs propres de la matrices Hessiennes de f soient > 0. (On dit que $H_f(x_0)$ est* définie positive*).*

- *Pour que x_0 soit un point de maximum local il est nécessaire que tous les valeurs propres de la matrices Hessiennes de f soient ≤ 0. (On dit que $H_f(x_0)$ est* semi-définie négative*).*

- *Pour que x_0 soit un point de maximum local il est suffisant que tous les valeurs propres de la matrices Hessiennes de f soient < 0. (On dit que $H_f(x_0)$ est* définie négative*).*

Dém. Si x_0 est un point de minimum local, alors $\exists\,\delta > 0$ tel que pour tout $\|h\| < \delta$ on a (voir (6.2))

$$\tfrac{1}{2}\langle H_f(x_0)h, h\rangle + o(\|h\|^2) = f(x_0 + h) - f(x_0) \geq 0.$$

Mais alors, avec les notations du lemme précédent, appliqué à $A = H_f(x_0)$, on trouve, pour $\|h\| < \delta$,

$$F(h) + o(1) \geq 0.$$

Soit $\lambda_1 = \min_{h\neq 0} F(h) = F(v_1)$. Comme $F(v_1) = F(\alpha\, v_1)$ pour tout $\alpha > 0$, en prenant $\alpha \to 0$ on déduit de l'inégalité précédente

$$\lambda_1 = \lim_{\alpha\to 0^+} F(\alpha\, v_1) \geq 0.$$

Mais alors toutes les valeurs propres de $H_f(x_0)$ sont ≥ 0. Cela prouve la première conclusion.

Supposons maintenant que toutes les valeurs propres soient strictement positives, et donc $\lambda_1 = \min_{h\neq 0} F(h) > 0$. En utilisant que

$$\tfrac{1}{\|h\|^2}\big[f(x_0 + h) - f(x_0)\big] = F(h) + o(1) \geq \lambda_1 + o(1), \qquad \text{pour } h \to 0.$$

on voit que pour $\|h\|$ assez petit cette expression est (strictement) positive. Donc x_0 est bien un point de minimum local (strict) pour f.

Les deux conclusions restantes se démontrent de la même manière. □

Remarque 6.7 (Points de selle)**.** Il arrive parfois qu'un point critique ne soit ni de minimum, ni de maximum local. C'est le cas, par exemple quand la plus petite et la plus grande valeur propre de la matrice Hessienne vérifient $\lambda_1 < 0 < \lambda_n$. On dit alors que x_0 est un *point de selle* (ou *de col*, ou de *min-max*). Par exemple, pour les fonctions de 2 variables, si (x_0, y_0) est un point de selle, alors la fonction $x \mapsto f(x, y_0)$ aura un minimum local en x_0 et la fonction $y \mapsto f(x, y)$ un maximum local en y_0 (ou l'inverse).

Règles pratiques pour appliquer le théorème 6.6 Pour appliquer le théorème 6.6 il n'est pas indispensable de calculer les valeurs propres de $H_f(x_0)$, puisque seul le signe des valeurs propres joue un rôle. Pour une fonctions de n variables, ces valeurs propres sont les solutions de l'équation (6.1), qui est une équation de la forme

$$\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = 0, \tag{6.5}$$

où le polynôme à gauche est le polynôme caractéristique de la matrice $H_f(x_0)$.

Exercice 6.8. Démontrer que les racines de (6.5) (dont on sait qu'elles sont toutes réelles) :
(i) sont toutes strictement négatives si et seulement si tous les coefficients $a_1 > 0, \ldots a_n > 0$.
(ii) sont toutes strictement positives si et seulement si les coefficients vérifient $a_1 < 0$, $a_2 > 0$, $a_3 < 0$, etc.

Exemple 6.9. La fonction $f(x, y, z) = x^2 + 2y^2 + z^2 + xy - xz$ a comme seul point stationnaire l'origine. On calcule aisément les dérivées partielles secondes en 0. On trouve alors.

$$\det(H_f(0) - \lambda I) = \det \begin{pmatrix} 2-\lambda & 1 & -1 \\ 1 & 4-\lambda & 0 \\ -1 & 0 & 2-\lambda \end{pmatrix} = -(\lambda^3 - 8\lambda^2 + 18\lambda - 10) = 0.$$

Compte tenu du résultat de l'exercice précédent, les valeurs propres de la matrice Hessienne sont toutes strictement positives. L'origine est alors un point de minimum local. Cet analyse ne donne aucun renseignement sur la nature globale de ce point de minimum.

Exercice 6.10 (Une méthode pratique pour les fonction de 2 variables)**.** Soit $f : \mathbb{R}^2 \to \mathbb{R}$ de classe C^2. Il est standard de noter $r = \frac{\partial^2 f}{\partial x^2}$, $s = \frac{\partial^2 f}{\partial x \partial y}$ et $r = \frac{\partial^2 f}{\partial y^2}$. Calculer $\det \begin{pmatrix} r-\lambda & s \\ s & t-\lambda \end{pmatrix}$ et en déduire que, en un point (x_0, y_0) :

- Si $rt - s^2 > 0$ alors (x_0, y_0) est un extremum local pour $f(x, y)$: un minimiseur si $r > 0$ et un maximiseur si $r < 0$.
- Si $rt - s^2 < 0$ alors (x_0, y_0) est un point de selle.

6.4 Extrema liés. Optimisation sous contrainte

Dans la section précédente nous avons étudié les problèmes d'optimisation de type

$$\min\{f(x) : x \in U\} \qquad \text{et} \qquad \max\{f(x) : x \in U\},$$

où l'ensemble $U \subset \mathbb{R}^n$ était un ouvert.

Dans cette section on s'intéresse aux problèmes analogues d'optimisation sur des ensembles fermés, par exemple,

$$\min\{f(x) : x \in F\}, \qquad \max\{f(x) : x \in F\},$$

où F est un fermé de $\mathbb{R}^n$.

Nous considérons le cas particulier important où F est une surface de niveau d'une fonction continue $g : \mathbb{R}^n \to \mathbb{R}$ (ou une ligne de niveau si $n = 2$). Sans perte de généralité on pourra alors supposer que

$$F = \{x \in \mathbb{R}^n : g(x) = 0\}.$$

Pour simplifier la présentation nous supposerons que g est une fonction C^1. Commençons par établir une propriété géométrique importante de l'ensemble F.

Proposition 6.7 (admise)**.** *Soit F l'ensemble des zéros de la fonction g et x_0 un point tel que $g(x_0) = 0$, $\nabla g(x_0) \neq 0$. Alors le vecteur $\nabla g(x_0)$ est orthogonal à la surface F au point x_0.*

La démonstration générale de cette proposition repose sur le théorème des fonctions implicites (hors programme. Voir le cours d'analyse du semestre suivant). Les deux exemples suivants illustrent la validité du résultat de la proposition dans deux particuliers.

Exemple 6.11. Considérons la fonction $g(x, y, z) = x^2 + y^2 + z^2 - 1$. Dans ce cas l'ensemble F est la sphère unité de $\mathbb{R}^3$. Si (x_0, y_0, z_0) est un point de la sphère, alors le vecteur normal à la sphère en ce point est parallèle à $\nabla g(x_0, y_0, z_0) = 2(x_0, y_0, z_0)$.

Exemple 6.12. Considérons le cas où $g(x,y) = 0 \iff y = \phi(x)$, où $\phi\colon \mathbb{R} \to \mathbb{R}$, autrement dit on suppose que l'on puisse expliciter y en fonction de x à l'aide de la relation $g(x,y) = 0$. Alors la courbe F est le graphe de la fonction ϕ. Observons que dans ce cas, en dérivant la relation $g(x,\phi(x)) = 0$ par rapport à x on trouve $g_x(x,\phi(x)) + g_y(x,\phi(x))\phi'(x) = 0$. Au point (x_0,y_0) cette égalité donne $\phi'(x_0) = -g_x(x_0,y_0)/g_y(x_0,y_0)$.

Mais la droite tangente au graphe de ϕ en (x_0,y_0) est la droite $y = y_0 + \phi'(x_0)(x-x_0)$, que l'on peut réécrire $-\phi'(x_0)x + y = c$, avec $c = -\phi'(x_0)x_0 + y_0$. Cette droite est celle de direction orthogonale au vecteur $(-\phi'(x_0),1)$. Ce vecteur est alors orthogonal à la courbe F au point (x_0,y_0). Mais, en utilisant la formule pour $\phi'(x_0)$, on voit que ce vecteur est parallèle à $(g_x(x_0,y_0), g_y(x_0,y_0))$. On retrouve ainsi dans ce cas particulier le résultat de la proposition, à savoir que le vecteur $(g_x(x_0,y_0), g_y(x_0,y_0))$ est orthogonal à cette courbe en (x_0,y_0).

Revenons-en aux problèmes d'optimisation avec "contrainte égalité"

$$\min\{f(x)\colon g(x)=0\}, \qquad \text{et} \qquad \min\{f(x)\colon g(x)=0\},$$

où g est de classe C^1. La condition $\nabla g(x_0) \neq 0$ exprime qu'en x_0 la contrainte *n'est pas dégénérée*. Voici un exemple de contrainte dégénéré : $xy = 0$ est une contrainte dégénéré à l'origine (observer que cette contrainte n'est pas une courbe mais plutôt l'intersection de deux courbes) à l'origine.

S'il n'y a avait pas la contrainte $g(x) = 0$, les solutions de ces problèmes d'optimisation seraient à chercher parmi les points stationnaires de la fonction f. Dans le cas d'une optimisation avec contrainte, la situation est différente :

Théorème 6.8 (des multiplicateurs de Lagrange. Admis)**.** *Si x_0 est un point tel que $g(x_0) = 0$ et $\nabla g(x_0) \neq 0$ et x_0 est un minimiseur ou un maximiseur pour $f(x)$ sous la contrainte $g(x) = 0$, alors $\nabla f(x_0) = 0$ est parallèle à $\nabla g(x_0)$. Autrement dit, il existe $\lambda_0 \in \mathbb{R}$ tel que $\nabla f(x_0) + \lambda_0 \nabla g(x_0) = 0$.*

On appelle le réel λ_0 le *multiplicateur de Lagrange* associé au problème d'optimisation. Le théorème précédent admet la reformulation suivante : introduisons la Lagrangienne du problème d'optimisation, qui est la fonction de $n+1$ variables

$$\mathcal{L}(x,\lambda) = f(x) + \lambda g(x).$$

On a

$$\begin{cases} \nabla f(x_0) + \lambda_0 \nabla g(x_0) = 0 \\ g(x_0) = 0 \end{cases} \iff \nabla \mathcal{L}(x_0,\lambda_0) = 0,$$

où $\nabla\mathcal{L}$ est le vecteur de $n+1$ composantes $\frac{\partial \mathcal{L}}{\partial x_1}, \dots \frac{\partial \mathcal{L}}{\partial x_n}$ et $\frac{\partial \mathcal{L}}{\partial \lambda}$.

Autrement dit, les points de minimum et maximum extrema de $f(x)$ sous la contrante $g(x) = 0$, sont à chercher parmi les points x_0 tels que (x_0,λ_0) est un point stationnaire de la Lagrangienne, ou éventuellement parmi les points où la contrainte $g(x) = 0$ est dégénérée : il s'agit alors de chercher les solutions de

$$\nabla\, \mathcal{L}(x,\lambda) = 0, \qquad \text{ou} \qquad \begin{cases} \nabla\, g(x) = 0 \\ g(x) = 0. \end{cases}$$

Observons que le théorème des multiplicateurs de Lagrange présenté ici ne donne qu'une condition nécessaire. Il est très utile pour trouver les points x_0 qui sont les bons candidats à être les solutions des problèmes d'optimisation avec contrainte. Mais il peut arriver que (x_0,λ_0) soit un point stationnaire de $\mathcal{L}$, ou que $g(x_0) = 0$ et $\nabla g(x_0) = 0$, sans que x_0 soit ni un minimum, ni un maximum du problème d'optimisation.

Optimimisation sous plusieurs contraintes. Le théorème des multiplicateurs de Lagrange se généralise au cas où il y a plusieurs contraintes à satisfaire. Considérons par exemple le problème

$$\min\{f(x)\colon g_1(x)=0,\ g_2(x)=0\}, \qquad \text{ou} \qquad \max\{f(x)\colon g_1(x)=0,\ g_2(x)=0\},$$

où f, g_1 et g_2 sont des fonctions de classe C^1. On introduit dans ce cas la Lagrangienne de $(n+2)$-variables $\mathcal{L}(x,\lambda_1,\lambda_2)$. On peut démontrer que les points de minimum ou maximum de f sous les contraintes $g_1(x)=g_2(x)=0$ sont à chercher parmi les points x_0 tels que $(x_0,(\lambda_0)_1,(\lambda_0)_2)$ est un un point stationnaire la Lagrangienne, ce qui conduit à étudier le système

$$\nabla\mathcal{L}(x,\lambda_1,\lambda_2)=0,$$

ou sinon parmi les points où au moins l'une des contraintes est dégénérée : ce-derniers sont les solutions de

$$\begin{cases}\nabla g_1(x)=0\\ g_1(x)=0\end{cases} \qquad \text{ou} \qquad \begin{cases}\nabla g_2(x)=0\\ g_2(x)=0\end{cases}$$

Bien entendu, ces considérations se généralisent à un nombre arbitraire de contraintes.