

Fonctions de plusieurs variables

1 Notions de dérivée

1.1 Prologue

Avant d'expliquer les notions de dérivées pour les fonctions de plusieurs variables, il est utile de se rappeler comment on procède pour définir la dérivée d'une fonction d'une variable. Soit $]a, b[$ un intervalle de \mathbb{R} , $f :]a, b[\rightarrow \mathbb{R}$ une fonction continue et $x_0 \in]a, b[$. Une première façon de dire que f est dérivable en x_0 consiste à regarder le taux de variation

$$\frac{f(x_0 + t) - f(x_0)}{t}, \quad \text{pour } t \neq 0 \text{ et } x_0 + t \in]a, b[,$$

et à demander que ce rapport admette une limite lorsque t tend vers 0. Nous ferons référence à ce point de vue comme étant celui de *Newton–Leibniz*, ces deux illustres savants en étant à l'origine.

Il existe un autre procédé, plus géométrique. Nous dessinons le graphe $\Gamma_f := \{(x, f(x)) \in]a, b[\times \mathbb{R} \mid x \in]a, b[\}$ et pour tout $x \in]a, b[$ différent de x_0 , nous traçons la droite Δ_x passant par les deux points $(x_0, f(x_0))$ et $(x, f(x))$. Lorsque l'on fait tendre x vers x_0 , on demande que

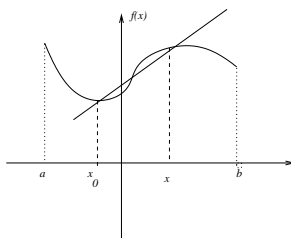


FIG. 1 – La droite Δ_x passant par les deux points $(x_0, f(x_0))$ et $(x, f(x))$ et le graphe de f

la droite Δ_x se positionne asymptotiquement vers une limite Δ_{x_0} , qui sera visualisée géométriquement comme la droite tangente à Γ_f au point $(x_0, f(x_0))$. Nous ferons référence à ce point

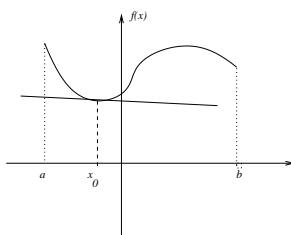


FIG. 2 – La droite limite Δ_{x_0} est la tangente à Γ_f au point $(x_0, f(x_0))$

de vue comme étant celui de *Fermat*. Notons que l'on peut qualifier la droite tangente en disant que c'est *la droite qui approche le mieux le graphe de f au voisinage du point $(x_0, f(x_0))$* .

La dérivabilité de f en x_0 se formulera en disant que :

- le taux de variation $\frac{f(x_0+t)-f(x_0)}{t}$ admet une limite, que l'on notera $f'(x_0)$ et que l'on appellera la dérivée de f en x_0 , si l'on adopte le point de vue de Newton–Leibniz,
 - la droite Δ_x admet une limite Δ_{x_0} lorsque x tend vers x_0 , que l'on appellera la droite tangente au graphe de f au point $(x_0, f(x_0))$, si l'on adopte le point de vue de Fermat.
- On fait le lien entre les deux points de vue en remarquant que $\frac{f(x_0+t)-f(x_0)}{t}$ est la *pente* de la droite Δ_x et sa limite $f'(x_0)$ est la pente de la droite tangente Δ_{x_0} .

Nous allons voir qu'essentiellement, si on cherche à transposer ces deux points de vue à des fonction à valeurs réelles de plusieurs variables, on obtient deux définitions différentes.

1.2 Dérivation selon un vecteur

On se place dorénavant dans \mathbb{R}^n muni des normes $\|\cdot\|_2, \|\cdot\|_\infty$, etc. (noter que, grâce aux résultats obtenus au chapitre précédent, on sait que le choix de la norme est indifférent pour tout ce qui concerne les notions de limite). On note (e_1, \dots, e_n) la base canonique de \mathbb{R}^n .

Soit U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ une fonction, $a \in U$ et $v \in \mathbb{R}^n$ un vecteur. Comme U est ouvert et $a \in U$, il existe $r > 0$ tel que la boule ouverte $B_2(a, r) := \{x \in \mathbb{R}^n \mid \|x - a\|_2 < r\}$ soit incluse dans U . En particulier, pour tout $t \in]-\frac{r}{\|v\|_2}, \frac{r}{\|v\|_2}[$, on a :

$$\|tv\|_2 < r \iff a + tv \in B_2(a, r) \implies a + tv \in U.$$

Ainsi l'application

$$\begin{aligned}]-\frac{r}{\|v\|_2}, \frac{r}{\|v\|_2}[&\longrightarrow \mathbb{R} \\ t &\longmapsto f(a + tv) \end{aligned}$$

est bien définie.

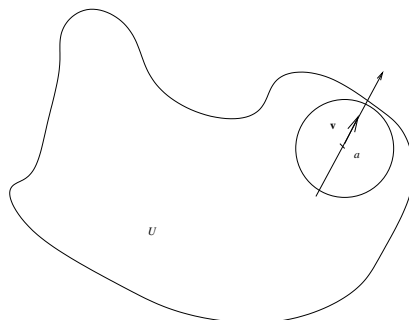


FIG. 3 – Si $t \in]-\frac{r}{\|v\|_2}, \frac{r}{\|v\|_2}[$, alors $a + tv \in U$

Définition 1 Soit U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ une fonction, $a \in U$ et $v \in \mathbb{R}^n$ un vecteur. On dit que « f est dérivable en a dans la direction v » ssi la fonction $t \mapsto f(a + tv)$ est dérivable en 0. Alors on note

$$D_v f(a) := \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t} \quad (1)$$

et on appelle cette quantité la dérivée de f dans la direction v en a .

Remarque 1 — Cette notion n'a d'intérêt que si $v \neq 0$. Par ailleurs si v et w sont deux vecteurs non nuls et colinéaires, c'est à dire, s'il existe $\lambda \in \mathbb{R}^*$ tel que $w = \lambda v$, alors

$$\frac{f(a + tw) - f(a)}{t} = \frac{f(a + t\lambda v) - f(a)}{t} = \lambda \frac{f(a + t\lambda v) - f(a)}{\lambda t} = \lambda \frac{f(a + sv) - f(a)}{s},$$

où $s := \lambda t$. Et donc on voit que $\frac{f(a+tw)-f(a)}{t}$ admet une limite lorsque $t \rightarrow 0$ ssi $\frac{f(a+sv)-f(a)}{s}$ admet une limite lorsque $s \rightarrow 0$. Donc « f est dérivable en a dans la direction v » ssi « f est dérivable en a dans la direction w ». Enfin en passant à la limite dans l'identité ci-dessus, on obtient que :

$$D_{\lambda v}f(a) = D_w f(a) = \lambda D_v f(a).$$

Remarque 2 — En pratique, nous n'utiliserons que des dérivés dans les directions e_1, \dots, e_n , où (e_1, \dots, e_n) est la base canonique de \mathbb{R}^n . Nous utilisons alors une notation spéciale pour désigner $D_{e_k}f(a)$: on note

$$\frac{\partial f}{\partial x_k}(a) := D_{e_k}f(a) := \lim_{t \rightarrow 0} \frac{f(a + te_k) - f(a)}{t}.$$

On appellera $\frac{\partial f}{\partial x_k}$ la « dérivée partielle de f par rapport à la variable x_k ». Analysons le sens de cette limite. Soit (x_1, \dots, x_n) les coordonnées de a dans la base (e_1, \dots, e_n) . Alors les coordonnées de $a + te_k$ sont :

$$(x_1, \dots, x_{k-1}, x_k + t, x_{k+1}, \dots, x_n).$$

Ainsi, pour calculer $\frac{\partial f}{\partial x_k}(a)$, on calcule la limite

$$\lim_{t \rightarrow 0} \frac{f(x_1, \dots, x_k + t, \dots, x_n) - f(x_1, \dots, x_n)}{t},$$

c'est à dire : **on gèle toutes les variables x_j , pour $j \neq k$, et on dérive par rapport à x_k** . Autrement dit, on se ramène à la dérivation d'une fonction d'une variable !

Exemple — Prenons la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = x^2 \cos y$$

et cherchons sa dérivée partielle par rapport à x pour toute valeur de (x, y) . Pour cela on gèle y (qui joue donc momentanément le rôle d'un paramètre) et on dérive par rapport à x . Cela donne :

$$\frac{\partial f}{\partial x}(x, y) = 2x \cos y.$$

De même, si on veut calculer la dérivée partielle de f par rapport à y , on gèle la variable x et on dérive par rapport à y :

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

Remarque 3 — Enfin nous pouvons observer que la définition de la dérivée que nous venons de voir est une généralisation aux fonctions de plusieurs variables du concept de dérivée selon Newton–Leibniz.

Définition 2 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction

- Si f admet une dérivée dans la direction v en tout point a de U , on dit alors que : « f admet une dérivée dans la direction v sur U »
- Si pour tout $k \in \llbracket 1, n \rrbracket$, f admet une dérivée dans la direction e_k sur U et si toutes les fonctions

$$\begin{array}{ccc} \frac{\partial f}{\partial x_k} : & U & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \frac{\partial f}{\partial x_k}(x) \end{array}$$

sont continues, on dit que : « f est de classe \mathcal{C}^1 sur U ».

1.3 Différentielle d'une fonction de plusieurs variables

L'idée est à présent de s'inspirer du point de vue de Fermat : la dérivée doit contenir l'information qui permet de trouver la meilleure approximation du graphe de f au voisinage d'un point $(a, f(a))$ qui soit un hyperplan. En effet, nous notons que, si f est une fonction d'un ouvert U de \mathbb{R}^n vers \mathbb{R} , alors son graphe $\Gamma_f := \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R} \mid x \in U\}$ est une hypersurface de \mathbb{R}^{n+1} . Au voisinage d'un point $(a, f(a))$, il est donc normal d'essayer d'approcher Γ_f par un hyperplan passant par $(a, f(a))$. Cet hyperplan peut être lui-même construit en prenant le graphe d'une fonction affine

$$F(x) = \alpha + \ell(x), \quad \text{où } \ell : \mathbb{R}^n \longrightarrow \mathbb{R} \text{ est linéaire.}$$

Le plus difficile dans l'histoire consiste à trouver la meilleure forme linéaire ℓ . Car, une fois que l'on a fixé ℓ , on en déduit facilement α : pour cela on demande que Γ_F passe le point $(a, f(a))$ ¹ et donc que $f(a) = F(a)$, ce qui entraîne $\alpha = f(a) - \ell(a)$ et donc $F(x) = f(a) - \ell(a) + \ell(x) = f(a) + \ell(x - a)$.

Supposons donc que α soit tel que $f(a) = F(a)$. On va choisir ℓ de façon à ce que $f(x)$ soit très très proche de $F(x)$ lorsque x est très proche de a . De façon plus précise, il est raisonnable de demander que le rapport

$$\frac{f(x) - F(x)}{x - a} \quad \text{tende vers 0 lorsque } x \rightarrow a.$$

Puisque $F(x) = f(a) + \ell(x - a)$, cela signifie que :

$$\frac{f(x) - f(a) - \ell(x - a)}{x - a} \quad \text{tende vers 0 lorsque } x \rightarrow a.$$

Définition 3 Soit U un ouvert de \mathbb{R}^n , $f : U \longrightarrow \mathbb{R}$ une fonction et $a \in U$. On dit que « f est différentiable en a » ssi il existe une application linéaire $\ell : \mathbb{R}^n \longrightarrow \mathbb{R}$ telle que

$$\lim_{h \in B(0, r); h \rightarrow 0} \frac{f(a + h) - f(a) - \ell(h)}{h} = 0. \quad (2)$$

Ou encore :

$$\forall a + h \in U, \quad f(a + h) = f(a) + \ell(h) + \|h\|\varepsilon(h),$$

où $\|\cdot\|$ est une norme (quelconque) et $\varepsilon(h)$ est une fonction qui s'annule en 0 et qui est continue en 0 (donc en particulier $\lim_{h \rightarrow 0} \varepsilon(h) = 0$). La forme linéaire ℓ est alors unique, est appelée « la différentielle de f en a » et est notée

$$df_a := \ell.$$

Remarque 1 — Une des différences avec la définition de la dérivabilité dans la direction d'un vecteur est que la limite dans (1) était la limite d'une fonction définie sur \mathbb{R} , tandis que la limite dans (2) est la limite d'une fonction définie sur un ouvert de \mathbb{R}^n et donc nécessite les notions de topologies vues au chapitre précédent pour être définie correctement.

Remarque 2 — Ainsi, si f admet une différentielle df_a en a , alors on a :

$$\forall a + h \in U, \quad f(a + h) = f(a) + df_a(h) + \|h\|\varepsilon(h), \quad \text{où } \lim_{h \rightarrow 0} \varepsilon(h) = 0.$$

Exemples de fonctions différentiable

¹c'est la moindre des choses si on demande que le graphe Γ_F de F approche Γ_f au voisinage du point $(a, f(a))$

a) **Les fonctions affines.** Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction affine, c'est à dire de la forme

$$f(x) = \alpha + \ell(x), \quad \text{où } \alpha \in \mathbb{R} \text{ et } \ell \in (\mathbb{R}^n)^*.$$

Alors, pour tout $a \in \mathbb{R}^n$,

$$f(a+h) = \alpha + \ell(a+h) = \alpha + \ell(a) + \ell(h) = f(a) + \ell(h)$$

et ℓ est linéaire. Donc f admet une différentielle en a , qui est ℓ ; i.e. $df_a = \ell$. Ainsi l'application $df : \mathbb{R}^n \rightarrow (\mathbb{R}^n)^*$ est constante et est égale à ℓ partout.

b) **La somme de deux fonctions différentiables.** Soit $U \subset \mathbb{R}^n$ un ouvert et f et g deux applications différentiables de U vers \mathbb{R} . Alors la somme

$$\begin{aligned} f+g : U &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) + g(x) \end{aligned}$$

est différentiable sur U et, $\forall a \in U$,

$$d(f+g)_a = df_a + dg_a.$$

La preuve est immédiate et est laissée au lecteur à titre d'exercice.

c) **Le produit de deux fonctions différentiables.** Soit $U \subset \mathbb{R}^n$ un ouvert et f et g deux applications différentiables de U vers \mathbb{R} . Alors le produit

$$\begin{aligned} fg : U &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x)g(x) \end{aligned}$$

est différentiable sur U et, $\forall a \in U$,

$$d(fg)_a = f(a)dg_a + g(a)df_a.$$

En effet nous avons, $\forall a \in U$,

$$f(a+h) = f(a) + df_a(h) + \|h\|\varepsilon_1(h) \quad \text{et} \quad g(a+h) = g(a) + dg_a(h) + \|h\|\varepsilon_2(h).$$

et en multipliant ces deux identités entre elles :

$$\begin{aligned} f(a+h)g(a+h) &= f(a)g(a) + f(a)dg_a(h) + g(a)df_a(h) \\ &\quad + [df_a(h)dg_a(h) + \|h\|(\varepsilon_1(h)(g(a) + dg_a(h)) + \varepsilon_2(h)(f(a) + df_a(h)))] , \end{aligned}$$

et on vérifie que le terme entre crochets est de la forme $\|h\|\varepsilon(h)$, où $\lim_{h \rightarrow 0} \varepsilon(h) = 0$.

d) **La composition d'une fonction différentiable avec une fonction dérivable.** Soit $U \subset \mathbb{R}^n$ un ouvert, $f : U \rightarrow \mathbb{R}$ une fonction différentiable, $]\alpha, \beta[$ un intervalle de \mathbb{R} et $g :]\alpha, \beta[\rightarrow \mathbb{R}$ une fonction dérivable. **On suppose que l'image $f(U)$ de f est contenue dans $]\alpha, \beta[$.** Alors

$$\begin{aligned} g \circ f : U &\longrightarrow \mathbb{R} \\ x &\longmapsto g(f(x)) \end{aligned}$$

est différentiable sur U et, $\forall a \in U$,

$$f(g \circ f)_a = g'(f(a)) df_a.$$

En effet nous avons, $\forall a \in U$,

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

et, pour $y \in \mathbb{R}$ tel que $f(a) + y \in]\alpha, \beta[$,

$$g(f(a) + y) = g(f(a)) + g'(f(a))y + |y|\theta(y).$$

Substituons $y = df_a(h) + \|h\|\varepsilon(h)$ dans cette dernière relation : nous obtenons

$$\begin{aligned} g \circ f(a+h) &= g(f(a) + df_a(h) + \|h\|\varepsilon(h)) \\ &= g(f(a)) + g'(f(a))(df_a(h) + \|h\|\varepsilon(h)) + |df_a(h) + \|h\|\varepsilon(h)|\theta(df_a(h) + \|h\|\varepsilon(h)) \\ &= g(f(a)) + g'(f(a))df_a(h) + \|h\|\varepsilon'(h), \end{aligned}$$

où l'on peut vérifier que

$$\varepsilon'(h) = g'(f(a))\varepsilon(h) + \frac{|df_a(h) + \|h\|\varepsilon(h)|}{\|h\|}\theta(df_a(h) + \|h\|\varepsilon(h))$$

tend vers 0 lorsque $h \rightarrow 0$. Donc $g \circ f$ est bien différentiable en a et $d(g \circ f)_a = g'(f(a))df_a$.

Exercice — A partir des exemples et des résultats précédents, démontrer que :

– tout polynôme

$$P(x) = \sum_{(k_1, \dots, k_n) \in [1, N]^n} a_{k_1 \dots k_n} (x_1)^{k_1} \dots (x_n)^{k_n}$$

de n variables réelles définit une fonction différentiable sur \mathbb{R}^n . Exprimer dP_x dans le cas où P est un polynôme de degré N égal à 2 (autrement dit, si P est une forme quadratique)

– toute fraction rationnelle $f = \frac{P}{Q}$ (où P et Q sont des polynômes de n variables réelles) définit une fonction différentiable sur $U := \{x \in \mathbb{R}^n \mid Q(x) \neq 0\}$.

– la fonction

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x, y) \longmapsto \frac{e^{x^2}}{1 + x^2 + y^2}$$

est différentiable sur \mathbb{R}^2 . Calculer sa différentielle en tout point $(x, y) \in \mathbb{R}^2$.

1.4 Lien entre les deux notions de dérivation

La chose la plus évidente est que la notion d'application différentiable est plus forte que celle de fonction dérivable selon un vecteur. C'est l'objet du résultat suivant.

Proposition 1 Soit U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ une fonction et $a \in U$. Si f est différentiable en a , alors pour tout vecteur $v \in \mathbb{R}^n$, f est dérivable en a dans la direction v et

$$D_v f(a) = df_a(v).$$

Démonstration — Supposons que f est différentiable en a . Cela nous donne en particulier que, pour tout $v \in \mathbb{R}^n$,

$$f(a+tv) = f(a) + df_a(tv) + \|tv\|\varepsilon(tv), \quad \text{où } \lim_{h \rightarrow 0} \varepsilon(h) = 0.$$

Nous utilisons cette relation pour écrire le taux de variations

$$\frac{f(a+tv) - f(a)}{t} = \frac{tdf_a(v) + |t| \cdot \|v\|\varepsilon(tv)}{t} = df_a(v) + \text{signe}(t)\varepsilon(tv).$$

Il est alors immédiat que $\frac{f(a+tv)-f(a)}{t}$ admet une limite lorsque t tend vers 0, qui est égale à $df_a(v)$.

Il est naturel de se demander si la réciproque est vraie. Là, les choses sont un peu plus compliquées. Il s'agit en effet de savoir si, étant donnée une fonction $f : U \rightarrow \mathbb{R}$ et $a \in U$, on peut déduire du fait que f est dérivable en a dans suffisamment de directions le fait que f est différentiable en a . D'abord il semble raisonnable de supposer que ce type de résultat n'ait lieu que si on sait que f est dérivable par rapport à au moins n vecteurs qui sont linéairement indépendants. Mais cela n'est en fait pas suffisant, comme le montre l'exemple qui suit.

Exemple — Nous considérons la fonction

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \frac{3x^2y - y^3}{x^2 + y^2}, \quad \text{si } (x, y) \neq 0$$

et nous posons $f(0, 0) = 0$, de sorte que f est continue sur \mathbb{R}^2 (exercice : vérifier !). Nous laissons au lecteur (encore à titre d'exercice) le soin de montrer que f est différentiable en tout point de $\mathbb{R}^2 \setminus \{(0, 0)\}$ et examinons ici ce qui se passe en $0 = (0, 0)$. Pour tout $\theta \in \mathbb{R}$, soit $v := (\cos \theta, \sin \theta)$. Alors pour tout $t \in \mathbb{R}^*$, on a

$$\frac{f(0 + tv) - f(0)}{t} = \frac{f(tv)}{t} = \frac{3t^3 \cos^2 \theta \sin \theta - t^3 \sin^3 \theta}{t(t^2 \cos^2 \theta + t^2 \sin^2 \theta)} = \frac{3 \cos^2 \theta \sin \theta - \sin^3 \theta}{\cos^2 \theta + \sin^2 \theta} = \sin(3\theta).$$

Nous voyons que cette quantité est indépendante de t , donc en particulier admet une limite lorsque $t \rightarrow 0$, égale à $\sin(3\theta)$. Or cette limite **n'est pas une fonction linéaire de v** , donc **f ne peut pas être différentiable en 0**. En effet supposons que f soit différentiable en 0. Alors, d'après la proposition précédente, on devrait avoir $\lim_{t \rightarrow 0} \frac{f(tv)}{t} = df_0(v)$, c'est à dire $3 \cos^2 \theta \sin \theta - \sin^3 \theta = df_a(\cos \theta, \sin \theta)$, ce qui est bien sûr impossible (puisque df_a est linéaire, on doit avoir $df_a(\cos \theta, \sin \theta) = \alpha \cos \theta + \beta \sin \theta$). Donc f n'est pas différentiable en 0.

Interprétation géométrique : le graphe de f est un cône de sommet $\{(0, 0, 0)\}$, c'est à dire une surface qui est la réunion d'une famille à un paramètre de demi-droites de $\mathbb{R}^2 \times \mathbb{R}$ qui passent toutes par l'origine. En particulier il n'y a pas de plan tangent au sommet du cône.

Nous allons voir maintenant, qu'avec des hypothèses plus fortes, nous avons une réciproque à la proposition précédente.

Théorème 1 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 , c'est à dire qui admet une dérivée $\frac{\partial f}{\partial x_k}(a)$ dans la direction e_k en a , pour tout $k \in \llbracket 1, n \rrbracket$ et pour tout $a \in U$, et telle que, $\forall k \in \llbracket 1, n \rrbracket$, $x \mapsto \frac{\partial f}{\partial x_k}(x)$ est continue sur U . Alors f est différentiable en chaque point de U . De plus on a, en tout point $a \in U$,

$$\forall x \in U, \quad df_a(x) = \sum_{k=1}^n D_{e_k} f(a) x_k = \sum_{k=1}^n \frac{\partial f}{\partial x_k}(a) x_k.$$

Démonstration — Pour simplifier la démonstration, nous ne donnons la preuve que pour le cas $m = 2$. L'idée est d'écrire, pour x_1 et x_2 petits,

$$f(a + (x_1, x_2)) - f(a) - x_1 \frac{\partial f}{\partial x_1}(a) - x_2 \frac{\partial f}{\partial x_2}(a) = \left[f(a + (x_1, x_2)) - f(a + (x_1, 0)) - x_2 \frac{\partial f}{\partial x_2}(a) \right] + \left[f(a + (x_1, 0)) - f(a) - x_1 \frac{\partial f}{\partial x_1}(a) \right]$$

et d'évaluer chacun des termes séparément. Par exemple pour le premier terme, nous observons que, puisque $D_{e_2}f$ existe partout, la fonction

$$t \longmapsto f(a + (x_1, tx_2))$$

est dérivable (et donc continue) sur $[-1, 1]$ et sa dérivée en t vaut $x_2 \frac{\partial f}{\partial x_2}(a + (x_1, tx_2))$. Donc nous pouvons lui appliquer le théorème des accroissements finis entre les valeurs 0 et 1 : $\exists \theta \in]0, 1[$ tel que

$$f(a + (x_1, x_2)) - f(a + (x_1, 0)) = x_2 \frac{\partial f}{\partial x_2}(a + (x_1, \theta x_2)),$$

En faisant de même avec $t \longmapsto f(a + (tx_1, 0))$, nous obtenons qu'il existe un réel $\tau \in]0, 1[$ tel que

$$f(a + (x_1, 0)) - f(a) = x_1 \frac{\partial f}{\partial x_1}(a + (\tau x_1, 0)).$$

Ainsi nous avons :

$$\begin{aligned} f(a + (x_1, x_2)) - f(a) - x_1 \frac{\partial f}{\partial x_1}(a) - x_2 \frac{\partial f}{\partial x_2}(a) &= x_2 \left(\frac{\partial f}{\partial x_2}(a + (x_1, \theta x_2)) - \frac{\partial f}{\partial x_2}(a) \right) \\ &\quad + x_1 \left(\frac{\partial f}{\partial x_1}(a + (\tau x_1, 0)) - \frac{\partial f}{\partial x_1}(a) \right). \end{aligned} \quad (3)$$

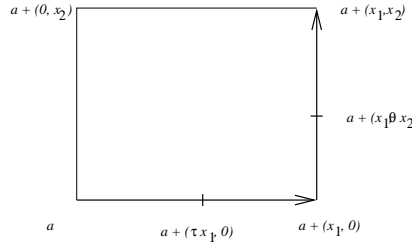


FIG. 4 –

Utilisons à présent le fait que $D_{e_1}f$ et $D_{e_2}f$ sont **continues** : pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que

$$\|x\|_\infty < \eta \implies \left\| \frac{\partial f}{\partial x_k}(a + x) - \frac{\partial f}{\partial x_k}(a) \right\| < \varepsilon.$$

Nous choisissons alors x tel que $\|x\|_\infty < \eta$ et lui appliquons l'identité (3). Cela entraîne (en remarquant qu'alors $\|(\tau x_1, 0)\|_\infty < \eta$ et $\|(x_1, \theta x_2)\|_\infty < \eta$) :

$$\begin{aligned} &\left| f(a + (x_1, x_2)) - f(a) - x_1 \frac{\partial f}{\partial x_1}(a) - x_2 \frac{\partial f}{\partial x_2}(a) \right| \\ &\leq \left| x_1 \left(\frac{\partial f}{\partial x_1}(a + (\tau x_1, 0)) - \frac{\partial f}{\partial x_1}(a) \right) + x_2 \left(\frac{\partial f}{\partial x_2}(a + (x_1, \theta x_2)) - \frac{\partial f}{\partial x_2}(a) \right) \right| \\ &\leq (|x_1| + |x_2|)\varepsilon. \end{aligned}$$

Donc

$$\lim_{\|x\| \rightarrow 0} \frac{|f(a + (x_1, x_2)) - f(a) - x_1 \frac{\partial f}{\partial x_1}(a) - x_2 \frac{\partial f}{\partial x_2}(a)|}{\|x\|} = 0.$$

Et cela prouve que f est différentiable en a .

1.5 Le théorème des accroissements finis

Pour étendre le théorème des accroissements finis au cas de plusieurs variables, nous avons besoin en premier lieu de trouver par quoi nous devons remplacer un intervalle de \mathbb{R} : par un sous-ensemble convexe de \mathbb{R}^n . D'abord, si a et b sont deux points de \mathbb{R}^n , nous définissons les intervalles

$$[a, b] := \{a + t(b - a) \mid t \in [0, 1]\} \subset \mathbb{R}^n \quad \text{et} \quad]a, b[:= \{a + t(b - a) \mid t \in]0, 1[\} \subset \mathbb{R}^n.$$

Puis nous dirons qu'un sous-ensemble $U \subset \mathbb{R}^n$ est **convexe** ssi $\forall a, b \in U$, on a $[a, b] \subset U$.

Théorème 2 Soit U un ouvert convexe de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 . Alors, pour tout $a, b \in U$, $\exists c \in]a, b[$ tel que

$$f(b) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(c)(b_i - a_i).$$

Démonstration — Soit $\varphi(t) := f(a + t(b - a)) - f(a) - (f(b) - f(a))t$. D'après les hypothèses, φ est une fonction \mathcal{C}^1 sur $[0, 1]$ et $\varphi(0) = \varphi(1) = 0$. Nous pouvons donc appliquer le théorème de Rolle à φ : $\exists \theta \in]0, 1[$ tel que $\varphi'(\theta) = 0$, ce qui est équivalent à :

$$\sum_{i=1}^n \frac{\partial f}{\partial x_i}(a + \theta(b - a)) = f(b) - f(a)$$

et cela nous donne le résultat avec $c = a + \theta(b - a)$.

1.6 Applications de classe \mathcal{C}^2

Soit $U \subset \mathbb{R}^n$ un ouvert et $f : U \rightarrow \mathbb{R}$ une fonction. Rappelons que f est \mathcal{C}^1 ssi f admet des dérivées partielles $\frac{\partial f}{\partial x_k}(x) := D_{e_k} f(x)$ en chaque point x de U et pour tout $k \in \llbracket 1, n \rrbracket$ et si, $\forall k \in \llbracket 1, n \rrbracket$, la fonction $\frac{\partial f}{\partial x_k} : U \rightarrow \mathbb{R}$ est continue.

Définition 4 On dit que la fonction $f : U \rightarrow \mathbb{R}$ est de classe \mathcal{C}^2 ssi f est de classe \mathcal{C}^1 , $\forall k \in \llbracket 1, n \rrbracket$, la fonction $\frac{\partial f}{\partial x_k} : U \rightarrow \mathbb{R}$ est différentiable et, pour tout $j, k \in \llbracket 1, n \rrbracket$, la fonction dérivée seconde partielle

$$\frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j} : U \rightarrow \mathbb{R}$$

est continue sur U .

On a alors le résultat suivant, appelé « lemme de Schwarz ».

Théorème 3 Soit $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 . Alors on a, $\forall j, k \in \llbracket 1, n \rrbracket$,

$$\forall a \in U, \quad \frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j}(a) = \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}(a).$$

Démonstration — Fixons $t, s \in \mathbb{R}^*$ tels que $a + te_j$ et $a + se_k$ soient dans la boule $B(a, r) \subset U$. Nous allons calculer de deux façons différentes la quantité

$$\mathcal{Q} := f(a + te_j + se_k) - f(a + te_j) - f(a + se_k) + f(a).$$

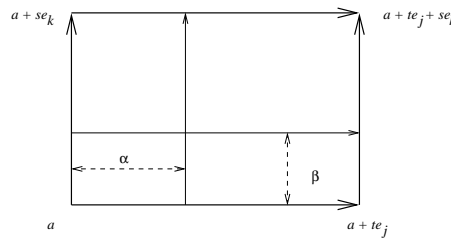


FIG. 5 – \mathcal{Q} est la somme des valeurs de f prises aux quatre sommets du rectangle avec des coefficients qui sont alternativement $+1$ et -1

1. Une famille continue et horizontale de sauts verticaux (cf. figure 1.6). Soit $\varphi(\alpha) := f(a + \alpha te_j + se_k) - f(a + \alpha te_j)$, $\forall \alpha \in [0, 1]$. Alors $\mathcal{Q} = \varphi(1) - \varphi(0)$. Comme f est de classe \mathcal{C}^1 , on peut appliquer une première fois la formule des accroissements finis : $\exists \theta_j \in]0, 1[$ tel que

$$\mathcal{Q} = \varphi(1) - \varphi(0) = \varphi'(\theta_j) = \frac{\partial f}{\partial x_j}(a + \theta_j te_j + se_k)t - \frac{\partial f}{\partial x_j}(a + \theta_j te_j)t.$$

Et comme f est de classe \mathcal{C}^2 on peut appliquer une deuxième fois le théorème des accroissements finis pour obtenir : $\exists \theta_k \in]0, 1[$ tel que

$$\mathcal{Q} = \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}(a + \theta_j te_j + \theta_k se_k)ts.$$

2. Une famille continue et verticale de sauts horizontaux. Soit $\psi(\beta) := f(a + te_j + \beta se_k) - f(a + \beta se_k)$, $\forall \beta \in [0, 1]$. Alors on a aussi $\mathcal{Q} = \psi(1) - \psi(0)$. En appliquant un raisonnement analogue, on obtient : $\exists \tau_k \in]0, 1[$ tel que

$$\mathcal{Q} = \psi(1) - \psi(0) = \psi'(\theta) = \frac{\partial f}{\partial x_k}(a + te_j + \tau_k se_k)s - \frac{\partial f}{\partial x_k}(a + \tau_k se_k)s.$$

Puis $\exists \tau_j \in]0, 1[$ tel que

$$\mathcal{Q} = \frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j}(a + \tau_j te_j + \tau_k se_k)ts.$$

On en déduit (en simplifiant par ts) que

$$\frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}(a + \theta_j te_j + \theta_k se_k) = \frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j}(a + \tau_j te_j + \tau_k se_k).$$

On fait alors tendre s et t vers 0 et on utilise le fait que $\frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}$ et $\frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j}$ sont continues. On obtient alors exactement la conclusion du théorème au point a .

Notation — Pour une fonction $f : U \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 , on notera désormais

$$\frac{\partial^2 f}{\partial x_j \partial x_k}(x) := \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_k}(x) = \frac{\partial \left(\frac{\partial f}{\partial x_k} \right)}{\partial x_j}(x).$$

Définition 5 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 . Pour tout point $x \in U$, la matrice hessienne de f est la matrice symétrique d'éléments $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$:

$$\text{Hess}(f)_x := \begin{pmatrix} \frac{\partial^2 f}{(\partial x_1)^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{(\partial x_n)^2}(x) \end{pmatrix}.$$

1.7 Formules de Taylor

Commençons par un rappel.

– **La formule de Taylor–Lagrange pour une fonction d'une variable réelle.** Soit $I \subset \mathbb{R}$ un intervalle et $f : I \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^{k+1} (c'est à dire qui est dérivable $k+1$ fois et dont la dérivée $(k+1)$ -ième $f^{(k+1)}$ est continue). Alors, si $[a, b] \subset I$, $\exists \theta \in]0, 1[$ tel que

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2}f''(a) + \cdots + \frac{(b-a)^k}{k!}f^{(k)}(a) + \frac{(b-a)^{k+1}}{(k+1)!}f^{(k+1)}(a + \theta(b-a)).$$

Démonstration — On part de la formule de Taylor avec reste intégral :

$$f(b) = \sum_{j=0}^k \frac{(b-a)^j}{j!} f^{(j)}(a) + \frac{(b-a)^{k+1}}{k!} \int_0^1 (1-t)^k f^{(k+1)}(a + t(b-a)) dt,$$

qui, rappelons-le, se démontre par récurrence sur k en faisant des intégrations par partie. Puis on cherche à exprimer le reste

$$R_k := \frac{(b-a)^{k+1}}{k!} \int_0^1 (1-t)^k f^{(k+1)}(a + t(b-a)) dt$$

différemment. Soit $m := \inf_{x \in [a, b]} f^{(k+1)}(x)$ et $M := \sup_{x \in [a, b]} f^{(k+1)}(x)$. Alors on a :

$$\forall t \in [0, 1], \quad m \leq f^{(k+1)}(a + t(b-a)) \leq M$$

et donc, en multipliant par $(1-t)^k$ et en intégrant sur $[0, 1]$,

$$\frac{m}{k+1} = \int_0^1 (1-t)^k m dt \leq \int_0^1 (1-t)^k f^{(k+1)}(a + t(b-a)) dt \leq \int_0^1 (1-t)^k M dt = \frac{M}{k+1},$$

ce qui donne, en multipliant par $k+1$:

$$m \leq \frac{(k+1)!}{(b-a)^{k+1}} R_k \leq M \iff \frac{(k+1)!}{(b-a)^{k+1}} R_k \in [m, M].$$

On utilise à présent le théorème des valeurs intermédiaires : puisque $f^{(k+1)}([a, b]) = [m, M]$, $\exists \theta \in]0, 1[$ tel que

$$f^{(k+1)}(a + \theta(b-a)) = \frac{(k+1)!}{(b-a)^{k+1}} R_k.$$

Cela nous donne la formule de Taylor–Lagrange annoncée plus haut.

Revenons à une fonction $f : U \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 , où U est un ouvert de \mathbb{R}^n .

Théorème 4 Soit $U \subset \mathbb{R}^n$ un ouvert, $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 et $a, b \in U$ deux points tels que $[a, b] \subset U$ (cela est vrai pour tous points a, b de U si U est convexe). Alors $\exists \theta \in]0, 1[$ tel que

$$f(b) = f(a) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a)(b_j - a_j) + \sum_{i,j=1}^n \frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(a + \theta(b - a))(b_i - a_i)(b_j - a_j).$$

Démonstration — Considérons la fonction

$$\begin{aligned} \varphi : [0, 1] &\longrightarrow \mathbb{R} \\ t &\longmapsto f(a + t(b - a)). \end{aligned}$$

On écrit la formule de Taylor–Lagrange pour φ à l'ordre 2 : $\exists \theta \in]0, 1[$ tel que

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{\varphi''(\theta)}{2}.$$

Puis, il ne reste plus qu'à calculer chaque terme :

$$\begin{aligned} \varphi(0) &= f(a), & \varphi(1) &= f(b) \\ \varphi'(0) &= \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a)(b_j - a_j), \\ \varphi''(t) &= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a + t(b - a))(b_i - a_i)(b_j - a_j), \end{aligned}$$

et le résultat est démontré.

Remarque — Nous pouvons encore écrire le développement sous la forme

$$f(b) = f(a) + df_a(b - a) + \frac{1}{2} Q_{a+\theta(b-a)}(b - a),$$

où

$$Q_x(\xi) := \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \xi_i \xi_j$$

est la forme quadratique sur \mathbb{R}^n dont la matrice dans la base canonique de \mathbb{R}^n est la matrice hessienne de f en x .

1.8 Points critique, points extrémaux

Définition 6 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 . On appelle **point critique de f** tout point $x \in U$ tel que $df_x = 0$. Si x est un point critique de f , le réel $f(x)$ est alors appelé **valeur critique de f** .

On peut formuler les choses différemment en disant qu'un point critique est un point $x \in U$ qui est solution du système de n équations

$$\frac{\partial f}{\partial x_1}(x) = \dots = \frac{\partial f}{\partial x_n}(x) = 0.$$

La notion de point critique est liée (mais **non identique**) à la notion suivante.

Définition 7 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction continue.

– un point $a \in U$ est appelé un **maximum local** de f ssi il existe une boule $B(a, r) \subset U$ telle que

$$\forall x \in B(a, r), \quad f(x) \leq f(a).$$

– un point $a \in U$ est appelé un **minimum local** de f ssi il existe une boule $B(a, r) \subset U$ telle que

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

D'une façon générale, un point qui est soit un maximum local, soit un minimum local est appelé un **extrémum local**. Si les inégalités précédentes ont lieu sur tout U (au lieu de $B(a, r)$) on parle alors de maximum global, minimum global ou d'extrémum global.

Un premier lien entre les deux notions (points critiques et extrémum local d'une fonction) est le suivant.

Proposition 2 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction différentiable partout. Soit $a \in U$ un extrémum local. Alors a est un point critique de f .

Démonstration — Nous raisonnons par l'absurde et supposons que a est un extrémum local (par exemple, sans perte de généralité, un maximum local), mais qu'en même temps $df_a \neq 0$. Cela signifie qu'il existe un vecteur $\xi \in \mathbb{R}^n$ tel que $df_a(\xi) \neq 0$. Alors ξ est forcément non nul et donc, en posant $v := \pm \xi / \|\xi\|$, on a encore $df_a(v) \neq 0$ et $\|v\| = 1$. De plus nous choisissons le signe \pm devant $\xi / \|\xi\|$ de façon à ce que l'on ait $df_a(v) > 0$. A présent nous écrivons que f est différentiable en a :

$$\forall a + h \in U, \quad f(a + h) = f(a) + df_a(h) + \|h\|\varepsilon(h), \quad \text{où } \lim_{h \rightarrow 0} \varepsilon(h) = 0$$

et nous exploitons cette identité avec $h = tv$, où $t \in \mathbb{R}$ est suffisamment proche de 0 pour que $a + tv \in U$. Cela donne

$$f(a + tv) = f(a) + tdf_a(v) + |t|\varepsilon(tv).$$

Comme $\lim_{h \rightarrow 0} \varepsilon(h) = 0$, il est possible de choisir $t > 0$ mais assez petit pour que $|\varepsilon(tv)| \leq \frac{1}{2}df_a(v)$. Alors

$$f(a + tv) \geq f(a) + tdf_a(v) - |t|\frac{1}{2}df_a(v) = f(a) + t\frac{1}{2}df_a(v) > f(a).$$

Et cela contredit le fait que a est un maximum local.

Remarque — Une hypothèse fondamentale dans ce résultat est que U **soit un ouvert**. En effet le résultat cesse d'être vrai en général sur un ensemble qui ne serait pas ouvert. Par exemple la fonction $f : [-1, 1] \rightarrow \mathbb{R}$ définie par $f(x) = x$ atteint son maximum en 1 et bien évidemment $f'(1) \neq 0$. La raison est que $[-1, 1]$ n'est pas un ouvert (c'est en l'occurrence un fermé).

En général, la réciproque au résultat précédent n'est pas vraie. Voici deux exemples qui illustrent cela.

- la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = x^3$ a un point d'inflexion en 0. En particulier 0 est un point critique de f , mais ça n'est ni un maximum, ni un minimum.
- la fonction

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x, y) \longmapsto x^2 - y^2$$

admet 0 comme unique point critique. Mais 0 n'est ni un ni un maximum, ni un minimum, c'est un *point selle* : quand on « regarde » le graphe de f d'une certaine façon, c'est à dire

si on étudie la restriction de f à la droite $\{(x, 0) \mid x \in \mathbb{R}\}$, 0 est alors un minimum local de cette restriction ; mais quand on « regarde » le graphe de f d'une autre façon, c'est à dire si on étudie la restriction de f à la droite $\{(0, y) \mid x \in \mathbb{R}\}$, 0 est alors un maximum local de cette restriction.

Théorème 5 Soit U un ouvert de \mathbb{R}^n et $f : U \longrightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 . Soit $a \in U$ un point tel que

- $df_a = 0$, i.e. a est un point critique de f
- la matrice hessienne de a définit une forme quadratique **définie positive**, i.e. $\forall(\xi_1, \dots, \xi_n) \in \mathbb{R}^n$,

$$\xi \longmapsto \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \xi_i \xi_j = \text{Hess}(f)_a(\xi) = Q_a(\xi)$$

est une forme définie positive.

Alors a est un **minimum local**.

De même, si

- $df_a = 0$, i.e. a est un point critique de f
- la matrice hessienne de a définit une forme quadratique **définie négative**.

Alors a est un **maximum local**.

Démonstration — Nous ne montrerons que le premier cas (si $\text{Hess}(f)_a$ est définie positive). Nous utiliserons la formule de Taylor donnée au théorème 4 :

$$f(b) = f(a) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a)(a_j - b_j) + \sum_{i,j=1}^n \frac{1}{2} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a + \theta(b - a)) \right) (b_i - a_i)(b_j - a_j).$$

Elle entraîne que, si l'on suppose que a est point critique,

$$f(b) = f(a) + \sum_{i,j=1}^n \frac{1}{2} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a + \theta(b - a)) \right) (b_i - a_i)(b_j - a_j) = f(a) + \frac{1}{2} Q_{a+\theta(b-a)}(b - a).$$

Et donc, si on est capable d'établir que :

$$\exists r > 0, \text{ tel que } \forall b \in B(a, r), \forall \theta \in [0, 1], \quad Q_{a+\theta(b-a)} > 0, \quad (4)$$

alors on aura montré que a est un minimum local. Nous allons consacrer le reste de la preuve à vérifier ce point délicat. Pour cela nous raisonnons par l'absurde et supposons le contraire de (4) :

$$\forall r > 0, \quad \exists c \in B(a, r) \text{ tel que : } Q_c \leq 0.$$

Nous choisissons $r_0 > 0$ tel que $B(a, r_0) \subset U$ et nous appliquons l'assertion précédente pour $r = \frac{r_0}{p+1}$, où $p \in \mathbb{N}$: pour chaque valeur de $p \in \mathbb{N}$, nous obtenons ainsi une valeur $c_p \in B(a, \frac{r_0}{p+1})$ telle que $Q_{c_p} < 0$. Cette dernière inégalité signifie qu'il existe un vecteur $\xi_p \in \mathbb{R}^n$ tel que

$$Q_{c_p}(\xi_p) \leq 0. \quad (5)$$

Sans perte de généralité, nous pouvons supposer que $\|\xi_p\| = 1, \forall p \in \mathbb{N}$. Nous avons ainsi une suite (c_p, ξ_p) qui prend ses valeurs dans le compact $\overline{B}(a, r_0) \times \overline{B}(0, 1)$ et telle que (5) ait lieu. Utilisons le théorème de Bolzano–Weierstrass : nous pouvons extraire une sous-suite $(c_{\varphi(p)}, \xi_{\varphi(p)})$ qui converge vers une limite $(c, \zeta) \in \overline{B}(a, r_0) \times \overline{B}(0, 1)$. Mais comme par ailleurs $\lim_{p \rightarrow \infty} c_p = a$, qui entraîne $\lim_{p \rightarrow \infty} c_{\varphi(p)} = a$, on en déduit que $c = a$. De plus la norme $\|\cdot\|$ sur \mathbb{R}^n étant une

fonction continue, nous avons $\|\zeta\| = \lim_{p \rightarrow \infty} \|\xi_{\varphi(p)}\| = \lim_{p \rightarrow \infty} 1 = 1$. A présent, nous utilisons simplement le fait que

$$(x, \xi) \longmapsto \text{Hess}(f)_x(\xi)$$

est une *fonction continue* sur $\overline{B}(a, r_0) \times \overline{B}(0, 1)$ et passons à la limite dans l'inégalité (5) (en y remplaçant p par $\varphi(p)$) : nous obtenons :

$$Q_a(\zeta) \leq 0, \quad (\text{mais en même temps } \|\zeta\| = 1!),$$

ce qui est bien entendu en contradiction avec l'hypothèse que Q_a est définie positive.

1.9 Etude en dimension 2

En guise d'applications des résultats précédents, voyons comment nous pouvons analyser une fonction de deux variables. En préliminaire voyons quelques propriétés des formes quadratiques en dimension deux. Soit Q une forme quadratique sur \mathbb{R}^2 . Dans les coordonnées (x, y) relatives à la base canonique, elle s'écrit :

$$Q(x, y) = px^2 + 2rxy + qy^2.$$

Et alors sa matrice dans la base canonique est

$$M = \begin{pmatrix} p & r \\ r & q \end{pmatrix}.$$

Nous savons que cette matrice symétrique est diagonalisable dans une base orthonormée, avec des valeurs propres réelles λ, μ . Il est intéressant de savoir retrouver les signes de λ et μ sans avoir à les calculer. Premièrement, du fait que

$$pq - r^2 = \det M = \lambda\mu,$$

on déduit que :

- la forme quadratique Q est non dégénérée ssi $pq - r^2 \neq 0$; alors les valeurs propres λ et μ sont non nulles et, en particulier, ont chacune un *signe bien défini*
- si $pq - r^2 > 0$, alors λ et μ sont de même signe
- au contraire, si $pq - r^2 < 0$, alors λ et μ sont de signe contraire.

Par ailleurs,

$$p + q = \text{tr}M = \lambda + \mu,$$

donc, **dans le cas où $pq - r^2 > 0$, c'est à dire si λ et μ sont de même signe,**

- si $p + q > 0$, alors λ et μ sont toutes les deux strictement positives
- si $p + q < 0$, alors λ et μ sont toutes les deux strictement négatives.

A présent, soit U un ouvert de \mathbb{R}^2 et $f : U \longrightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 et posons-nous la question de savoir où sont les extréma locaux de f (s'ils existent).

- D'abord, nous savons que, s'il y a un extrémum, il fait partie de l'ensemble des points critiques. La première tâche consiste donc à rechercher tous les points critiques de f , c'est à dire les solutions (x, y) de l'équation

$$df_{(x,y)} = 0 \iff \frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0.$$

- Ensuite, pour chaque point critique a , il faut se demander s'il est un extrémum. Pour cela, on commence par calculer la matrice hessienne de f en a (c'est à dire la matrice de la forme quadratique hessienne Q_a). Notons $p = \frac{\partial^2 f}{(\partial x)^2}(a)$, $q = \frac{\partial^2 f}{(\partial y)^2}(a)$ et $r = \frac{\partial^2 f}{\partial x \partial y}(a)$.

- Si $pq - r^2 \neq 0$ la forme quadratique est non dégénérée.
- Si $pq - r^2 > 0$ et $p + q > 0$, les valeurs propres de la matrice hessienne sont toutes les deux strictement positives, donc Q_a est définie positive, donc, d'après le théorème 5, a est un *mimum local*.
- Si $pq - r^2 > 0$ et $p + q < 0$, les valeurs propres de la matrice hessienne sont toutes les deux strictement négatives, donc Q_a est définie négative, donc, d'après le théorème 5, a est un *maximum local*.
- Si $pq - r^2 < 0$ les deux valeurs propres de la forme quadratique sont de signes contraire, le point critique a est un **point selle** (en particulier, il n'est pas un extrémum local).
- Si $pq - r^2 = 0$, la forme quadratique Q_a est dégénérée, **on ne peut rien conclure en général**.