Résumé de calcul différentiel

Thomas Letendre

PCSI - UE Math2 Printemps 2013

1 Fonctions, limites, continuité

Définition 1.1. Une *application* f est la donnée d'un ensemble de départ U, d'un ensemble d'arrivée V, et pour tout point $x \in U$ la donnée de son image $f(x) \in V$.

Définition 1.2. Soient $f : U \to V$ et $g : V' \to W$ deux applications avec $V \subset V'$. On définit *l'application composée* $g \circ f : U \to W$ par : $\forall x \in U, (g \circ f)(x) = g(f(x))$.

Remarques 1.3.

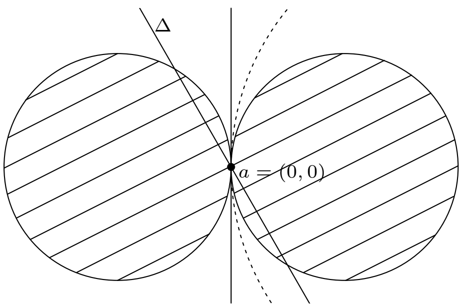
1. Ici on aura toujours $U \subset \mathbb{R}^n$ et $V = \mathbb{R}^m$. On parle alors de *fonction vectorielle*. On parle de *fonction* lorsque $m = 1$ ou, par abus de langage, pour m quelconque si le contexte est clair.
2. La formule définissant $g \circ f$ n'a du sens que si $V \subset V'$, c'est-à-dire si on peut écrire un diagramme : $U \xrightarrow{f} V \xrightarrow{g} W$ où l'ensemble intermédiaire est à la fois l'espace but de f et contenu dans l'espace source de g.

Notations 1.4. Pour tout $x = (x_1, \dots, x_n)$ et $y = (y_1 \dots, y_n) \in \mathbb{R}^n$, on note $x \cdot y = \sum_{j=1}^n x_j y_j$ le *produit scalaire* de x et y, et $\|x\| = \sqrt{x \cdot x} = (\sum_{j=1}^n x_j^2)^{\frac{1}{2}}$ la *norme* de x.

Définition 1.5. Soient $f : U \to \mathbb{R}^m$, $l \in \mathbb{R}^m$ et $a \in \overline{U}$, où $\overline{U}$ désigne $U \cup \partial U$. On dit que f tend vers l en a, et on note $f(x) \xrightarrow[x \to a]{} l$, lorsque : $\forall \varepsilon > 0\ \exists \alpha > 0\ ((x \in U \text{ et } \|x - a\| < \alpha) \Rightarrow (\|f(x) - l\| < \varepsilon))$.

Existence d'une limite. Souvent, pour prouver que $f(x) \xrightarrow[x \to a]{} l$, on majore $\|f(x) - l\|$ par une quantité qui ne dépend que de $\|x - a\|$ et tend vers 0 lorsque $\|x - a\|$ tend vers 0.
Si $l = (l_1, \dots, l_m) \in \mathbb{R}^m$ et $f = (f_1, \dots, f_m)$, où $f_i : U \to \mathbb{R}$ est la i-ème composante de f, on peut aussi montrer que pour tout $i \in \{1, \dots, m\}$, $f_i(x) \xrightarrow[x \to a]{} l_i$, ce qui est équivalent.

Remarque 1.6. Dire que f tend vers l en a est plus fort que dire que les n fonctions vectorielles partielles tendent vers l en a, ou même que la restriction de f à toute droite passant par a tend vers l en a.
Par exemple, considérons $f : \mathbb{R}^2 \to \mathbb{R}$ comme ci-dessous, qui vaut 1 sur l'axe vertical et dans les deux disques hachurés, et valant 0 partout ailleurs.



Une droite passant par $a = (0, 0)$ et qui n'est pas l'axe des ordonnées, par exemple Δ, contient un segment centré en a inclus dans la zone hachurée.

En restriction à une droite quelconque passant par a, f est donc constante à 1 sur un segment autour de a. En particulier, en restriction à toute droite passant par a, f tend vers 1 en a.

Cependant f ne tend pas vers 1 en a. En effet, on peut trouver des points arbitrairement proches de a en lesquels f s'annule, par exemple des points de la courbe en pointillés.

Proposition 1.7. *Soient f et $\tilde{f} : U \to \mathbb{R}^m$, $\mu : U \to \mathbb{R}$ et $a \in \overline{U}$ telles que $f(x) \xrightarrow[x \to a]{} l$, $\tilde{f}(x) \xrightarrow[x \to a]{} \tilde{l}$ et $\mu(x) \xrightarrow[x \to a]{} \lambda$. Alors $(f + \tilde{f})(x) \xrightarrow[x \to a]{} l + \tilde{l}$ et $(\mu f)(x) \xrightarrow[x \to a]{} \lambda l$.*

Proposition 1.8. *Soient $f : U \to V$ et $g : V \to \mathbb{R}^p$ avec $V \subset \mathbb{R}^m$. Soient $a \in \overline{U}$, $b \in \overline{V}$ et $l \in \mathbb{R}^p$, si $f(x) \xrightarrow[x \to a]{} b$ et $g(y) \xrightarrow[y \to b]{} l$, alors $(g \circ f)(x) \xrightarrow[x \to a]{} l$.*

Définition 1.9. Soient $f : U \to \mathbb{R}^m$ et $a \in U$, on dit que f est *continue* en a si $f(x) \xrightarrow[x \to a]{} f(a)$. On dit que f est *continue* (ou *de classe $\mathcal{C}^0$*) *sur* U, si elle est continue en tout point $a \in U$.

Remarques 1.10. 1. Si $f = (f_1, \dots, f_m)$, il est équivalent de dire que f est continue en a (resp. sur U) et de dire que toutes ses composantes (f_i) sont continues en a (resp. sur U).

2. En revanche, dire que f est continue en a est strictement plus fort que dire que les fonctions vectorielles partielles de f sont continues en a, ou même que la restriction de f à toute droite passant par a est continue en a. Le contre-exemple de la remarque 1.6 s'adapte immédiatement.

Proposition 1.11. *Soient f et $\tilde{f} : U \to \mathbb{R}^m$, $\mu : U \to \mathbb{R}$ et $a \in U$. Si f, $\tilde{f}$ et μ sont continues en a (resp. sur U), alors $f + \tilde{f}$ et μf sont continues en a (resp. sur U).*

Proposition 1.12. *Soient $f : U \to V$ et $g : V \to \mathbb{R}^p$ avec $V \subset \mathbb{R}^m$. Si f est continue en $a \in U$ (resp. sur U) et g est continue en $f(a) \in V$ (resp. sur V), alors $g \circ f$ est continue en a (resp. sur U).*

Exemples 1.13. Les fonctions coordonnées $(x_1, \dots, x_n) \mapsto x_j$ pour $j \in \{1, \dots, n\}$, les fonctions polynomiales (en les n variables $x_1, \dots, x_n$), les fractions rationnelles (quotients de deux polynômes), les fonctions trigonométriques cos, sin, ..., les fonctions exp et ln sont continues partout où elles sont définies.

Continuité d'une fonction vectorielle. Souvent, pour prouver qu'une fonction vectorielle est continue, on remarque que ses composantes sont obtenues par opérations élémentaires (sommes, produits, compositions) à partir de fonctions que l'on sait continues (par exemple celles de l'exemple 1.13) et sont donc continues par 1.11 et 1.12. Quand cela échoue, il faut revenir à la définition et calculer la limite de f aux points considérés (exemple : TD fiche 2, exercice 5.1).

2 Petits o

Définition 2.1. Soient f et $g : U \to \mathbb{R}$ et $a \in \overline{U}$, on dit que f est un *petit o* de g en a si il existe une fonction $\epsilon : U \to \mathbb{R}$ telle que $f = \epsilon g$ et $\epsilon(x) \xrightarrow[x \to a]{} 0$. On note $f(x) = \underset{x \to a}{o}(g(x))$.

Cela signifie que $f(x)$ est "négligeable" devant $g(x)$ au voisinage du point a. On omet souvent de préciser le point a dans la notation, mais il faut retenir qu'il s'agit d'une information ponctuelle.

Exemple 2.2. Sur $\mathbb{R}$, $x^2 = \underset{x \to 0}{o}(x)$, ce qui signifie que x^2 tend vers 0 "plus vite" que x en 0.

Pour $f : U \to \mathbb{R}$ et $a \in \overline{U}$, on a l'équivalence entre "$f(x) = o(1)$ en a" et $f(x) \xrightarrow[x \to a]{} 0$.
De même, "$f(x) = f(a) + o(1)$ en a" est équivalent à la continuité de f en a.

3 Rappels d'algèbre linéaire

Définition 3.1. Une application $L : \mathbb{R}^n \to \mathbb{R}^m$ est dite *linéaire*, si pour tout x et $y \in \mathbb{R}^n$ et tout $\lambda \in \mathbb{R}$ on a : $L(x + y) = L(x) + L(y)$ et $L(\lambda x) = \lambda L(x)$.

Remarques 3.2. 1. On note 0 les vecteurs $(0, \dots, 0)$ de $\mathbb{R}^n$ et $\mathbb{R}^m$. Pour L linéaire, on a $L(0) = 0$.

2. L est linéaire si et seulement si ses composantes $L_i : \mathbb{R}^n \to \mathbb{R}$ le sont pour $i \in \{1, \dots, m\}$.

Exemples 3.3. 1. $L : \mathbb{R} \to \mathbb{R}$ définie par $x \mapsto -5x$,

2. $\tilde{L} : (x, y, z) \mapsto (x + 2y, 3y - 6z)$ de $\mathbb{R}^3$ dans $\mathbb{R}^2$,

3. la projection de $\mathbb{R}^n$ sur la j-ème coordonnée, $(x_1, \dots, x_n) \mapsto x_j$.

Proposition 3.4. *Soient L et $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^m$ linéaires et $\lambda \in \mathbb{R}$, alors $L + \tilde{L}$ et λL sont linéaires.*

Proposition 3.5. *Soient $L : \mathbb{R}^n \to \mathbb{R}^m$ et $K : \mathbb{R}^m \to \mathbb{R}^p$ linéaires, alors $(K \circ L) : \mathbb{R}^n \to \mathbb{R}^p$ est linéaire.*

Notations 3.6. On note $e_1, \dots, e_n$ les vecteurs de la base canonique de $\mathbb{R}^n$ (les coordonnées de e_j sont toutes nulles sauf la j-ème qui vaut 1). Un point $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ s'écrit alors comme $x = \sum\limits_{j=1}^{n} x_j e_j$.

Soit $L : \mathbb{R}^n \to \mathbb{R}^m$ linéaire, pour tout $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $L(x) = L\left(\sum\limits_{j=1}^{n} x_j e_j\right) = \sum\limits_{j=1}^{n} x_j L(e_j)$. L'application linéaire L est donc totalement déterminée par la donnée de n vecteurs de $\mathbb{R}^m$: $L(e_1), \dots, L(e_n)$. Pour $m = 1$, cela montre que si $L : \mathbb{R}^n \to \mathbb{R}$ est linéaire elle est de la forme : $(x_1, \dots, x_n) \mapsto b_1 x_1 + \cdots + b_n x_n$ avec $b_1, \dots, b_n \in \mathbb{R}$. Il existe donc un unique $b = (b_1, \dots, b_n) \in \mathbb{R}^n$, tel que $\forall x \in \mathbb{R}^n, L(x) = b \cdot x$.

Définition 3.7. Une *matrice* de taille $m \times n$ est un tableau rectangulaire de réels, à m lignes et n colonnes.

Notations 3.8. On note $\mathcal{M}_{m,n}$ l'ensemble des matrices de taille $m \times n$. Si $M \in \mathcal{M}_{m,n}$, on note m_{ij} le coefficent réel figurant à l'intersection de la i-ème ligne et de la j-ème colonne de M, pour $1 \leqslant i \leqslant m$ et $1 \leqslant j \leqslant n$. On note aussi $M = (m_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$, où l'ensemble d'indexation est souvent omis.

Définition 3.9. Soient $A \in \mathcal{M}_{p,m}$ et $B \in \mathcal{M}_{m,n}$ avec $A = (a_{ij})$ et $B = (b_{ij})$. On définit le produit AB comme la matrice $(c_{ij}) \in \mathcal{M}_{p,n}$ telle que pour tout $i \in \{1, \ldots, p\}$ et tout $j \in \{1, \ldots, n\}$, $c_{ij} = \sum\limits_{k=1}^{m} a_{ik} b_{kj}$.

Remarque 3.10. Cette définition n'a de sens que si le nombre de colonnes de A est égal au nombre de lignes de B. Par ailleurs, le produit matriciel n'est pas commutatif. Ici BA n'a de sens que si $n = p$ et, même dans ce cas, AB et BA sont a priori de tailles différentes, donc ne peuvent être égales.

Définition 3.11. Soit $L : \mathbb{R}^n \to \mathbb{R}^m$ linéaire, on appelle *matrice de* L la matrice $\texttt{Mat}(L) \in \mathcal{M}_{m,n}$ dont la j-ème colonne contient les m coordonnées de $L(e_j)$:

$$\texttt{Mat}(L) = \begin{pmatrix} L(e_1)_1 & L(e_2)_1 & \cdots & L(e_n)_1 \\ L(e_1)_2 & L(e_2)_2 & \cdots & L(e_n)_2 \\ \vdots & \vdots & & \vdots \\ L(e_1)_m & L(e_2)_m & \ldots & L(e_n)_m \end{pmatrix}$$

où on a noté $L(e_j) = (L(e_j)_1, \ldots, L(e_j)_m)$ pour tout $j \in \{1, \ldots, n\}$.

Exemples 3.12. Pour les exemples de 3.3 1 et 2, on a $\texttt{Mat}(L) = \begin{pmatrix} -5 \end{pmatrix} \in \mathcal{M}_{1,1}$ et $\texttt{Mat}(\tilde{L}) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & -6 \end{pmatrix}$.

Dans la suite, on identifiera naturellement un vecteur $x \in \mathbb{R}^n$ avec la matrice colonne (de taille $n \times 1$) qui contient ses coordonnées. Alors, pour tout $L : \mathbb{R}^n \to \mathbb{R}^m$ linéaire et tout $x \in \mathbb{R}^n$,

$$L(x) = \begin{pmatrix} L(x)_1 \\ \vdots \\ L(x)_m \end{pmatrix} = \begin{pmatrix} \sum\limits_{j=1}^{n} x_j L(e_j)_1 \\ \vdots \\ \sum\limits_{j=1}^{n} x_j L(e_j)_m \end{pmatrix} = \begin{pmatrix} L(e_1)_1 & \cdots & L(e_n)_1 \\ \vdots & & \vdots \\ L(e_1)_m & \ldots & L(e_n)_m \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \texttt{Mat}(L) \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \texttt{Mat}(L)x.$$

Proposition 3.13. *Soient L et $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^m$ linéaires et $\lambda \in \mathbb{R}$, alors* $\texttt{Mat}(L + \tilde{L}) = \texttt{Mat}(L) + \texttt{Mat}(\tilde{L})$ *et* $\texttt{Mat}(\lambda L) = \lambda \texttt{Mat}(L)$ *(l'addition et la multiplication par un réel étant définies coefficient à coefficient).*

Proposition 3.14. *Soient $L : \mathbb{R}^n \to \mathbb{R}^m$ et $K : \mathbb{R}^m \to \mathbb{R}^p$ linéaires, alors* $\texttt{Mat}(K \circ L) = \texttt{Mat}(K)\texttt{Mat}(L)$.

Définition 3.15. Soit $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathcal{M}_{2,2}$, le *déterminant* de M est le réel $\det(M) = ad - bc$.

Remarque 3.16. La notion de déterminant se généralise à la dimension n, mais la formule est compliquée. Le déterminant n'a de sens que pour des matrices carrées.

4 Différentiabilité

Définition 4.1. Une partie $U \subset \mathbb{R}^n$ est dite *ouverte*, si pour tout $a \in U$, il existe $r > 0$ tel que la boule ouverte de centre a et de rayon r, $B(a, r) = \{x \in \mathbb{R}^n | \|x - a\| < r\}$, soit contenue dans U.

Dans la suite, les ensembles de définition de toutes les fonctions vectorielles considérées seront supposés ouverts. Par défaut, U désigne toujours une partie de $\mathbb{R}^n$.

4.1 Dérivées directionnelles et partielles

Définition 4.2. Soient $f : U \to \mathbb{R}^m$, $a \in U$ et $h \in \mathbb{R}^n$. Si $\frac{f(a+th)-f(a)}{t}$ a une limite quand t tend vers 0, on dit cette limite est la *dérivée directionnelle* de f en a dans la direction h, notée $[D_h f](a)$ ou $[\partial_h f](a)$.

Remarques 4.3. 1. Pour étudier la limite ci-dessus, il faut que f soit définie en $a + th$, au moins pour des t assez proches de 0. C'est pourquoi on travaille sur des ouverts lorsqu'on veut parler de dérivée (directionnelle, partielle, mais aussi classique pour $f : \mathbb{R} \to \mathbb{R}^m$) ou de différentielle (voir 4.2).

2. Par définition, f est dérivable dans la direction h en a si et seulement si $\varphi_h : t \mapsto f(a+th)$ est dérivable en 0. On a alors $\varphi'(0) = [D_h f](a) \in \mathbb{R}^m$.

Définition 4.4. Soit $f : U \to \mathbb{R}^m$, $a \in U$ et $e_1, \dots, e_n$ les vecteurs de la base canonique de $\mathbb{R}^n$. Si elle existe, la dérivée directionnelle $[D_{e_j} f](a)$ est appellée *dérivée partielle* de f en a, par rapport à la j-ème variable. On la note $\frac{\partial f}{\partial x_j}(a)$.

Notation 4.5. On note souvent $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$ si $n = 2$, et $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3})$ si $n = 3$.

4.2 Différentielle

Définition 4.6. Soient $f : U \to \mathbb{R}^m$ et $a \in U$, on dit que f est *différentiable* en a s'il existe une application linéaire $L : \mathbb{R}^n \to \mathbb{R}^m$ telle que $\|f(a+h) - f(a) - L(h)\| = \underset{h \to 0}{o}(\|h\|)$, c'est-à-dire :

$$\frac{f(a+h) - f(a) - L(h)}{\|h\|} \xrightarrow[h \to 0]{} 0.$$

L'application L est alors appelée la *différentielle* de f en a, et est notée $[df](a)$. On appelle $\mathtt{Mat}([df](a))$ la *matrice jacobienne* de f en a, notée $J(f)(a)$ ou $\mathtt{Jac}(f)(a)$. Pour tout $h \in \mathbb{R}^n$, on note $[df](a).h$ au lieu de $[df](a)(h)$.

Proposition 4.7. *Si $f : U \to \mathbb{R}^m$ est différentiable en a, alors sa différentielle est unique.*

Remarques 4.8.
1. Notons $f_1, \dots, f_m$ les composantes de f. Alors f est différentiable en a si et seulement si les f_i sont toutes différentiables en a, et $[df](a).h = ([df_1](a).h, \dots, [df_m](a).h)$ pour tout $h \in \mathbb{R}^n$.
2. La définition signifie que $x \mapsto f(a) + [df](a).(x-a)$ est la meilleure approximation de f par une fonction vectorielle affine (linéaire plus une constante) au voisinage de a.

Exemples 4.9.
1. Les fonctions de l'exemple 1.13 sont différentiables partout où elles sont définies.
2. Une application linéaire $L : \mathbb{R}^n \to \mathbb{R}^m$ est différentiable sur $\mathbb{R}^n$, et $\forall x \in \mathbb{R}^n$, $[dL](x) = L$. Par exemple, $(x_1, \dots, x_n) \mapsto x_j$ est égale à sa différentielle en un point quelconque. On la note souvent dx_j, car c'est la différentielle en un point quelconque de la j-ème fonction coordonnée.

Proposition 4.10. *Si $f : U \to \mathbb{R}^m$ est différentiable au point $a \in U$, alors elle est continue en a.*

Remarque 4.11. La réciproque est fausse, $x \mapsto |x|$ est continue sur $\mathbb{R}$, mais n'est pas différentiable en 0.

Proposition 4.12. *Si $f : U \to \mathbb{R}^m$ est différentiable en $a \in U$, alors f admet des dérivées directionnelles en a dans toutes les directions. De plus, pour tout $h \in \mathbb{R}^n$, on a $[df](a).h = [D_h f](a)$.*

Remarque 4.13. La réciproque est fausse. Un contre-exemple est fourni par la fonction vectorielle f de la remarque 1.6. Elle a une dérivée directionnelle dans toutes les directions en $(0,0)$, mais elle n'y est pas continue, donc pas différentiable. Voir aussi le TD, fiche 2 exercice 5.

Soit $f : U \to \mathbb{R}^m$ différentiable en $a \in U$, alors $\forall j \in \{1, \dots, n\}$, $[df](a).e_j = [D_{e_j} f](a) = \frac{\partial f}{\partial x_j}(a)$. Donc (voir la section 3), pour tout $h \in \mathbb{R}^n$, $[df](a).h = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a)$.

En tant qu'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$, $[df](a) = \frac{\partial f}{\partial x_j}(a) dx_j$, où dx_j est l'application définie dans l'exemple 4.9 2. On a aussi :

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

Existence de dérivées directionnelles ou partielles. Pour prouver que $f : U \to \mathbb{R}^m$ a une dérivée en a dans la direction h (en particulier une dérivée partielle), si f est différentiable en a on conclut par 4.12. Sinon on revient à la définition 4.2, en montrant l'existence d'une limite, ou que $\varphi_h : t \mapsto f(a+th)$ est dérivable en 0 (voir remarque 4.3 2) ce qui est souvent plus simple quand f est définie par une formule.

Calcul des dérivées partielles. Sous réserve d'existence, $\frac{\partial f}{\partial x_j}(a)$ est la dérivée au point $a_j \in \mathbb{R}$ de la fonction vectorielle partielle $x_j \mapsto f(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n)$ de $\mathbb{R}$ dans $\mathbb{R}^m$. On calcule donc $\frac{\partial f}{\partial x_j}$ en dérivant f par rapport à x_j, les autres variables étant traitées comme des constantes.

Calcul de la différentielle. Si elle existe, la différentielle de f en a est : $[df](a) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a)dx_j$. Il suffit donc de calculer les dérivées partielles de f en a. Il est strictement équivalent de calculer $J(f)(a)$.

Calcul d'une dérivée directionnelle. Si f est différentiable en a et $h \in \mathbb{R}^n$, on calcule $[D_h f](a)$ à partir de la différentielle ou des dérivées partielles par la relation : $[D_h f](a) = [df](a).h = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a)$. Matriciellement $[D_h f](a) = (J(f)(a))h$, en identifiant vecteurs et matrices colonnes associées. Si $[D_h f](a)$ existe mais f n'est pas différentiable en a, il n'y a pas de relations entre $[D_h f](a)$ et les éventuelles dérivées partielles de f en a. On utilise alors la relation $[D_h f](a) = \varphi_h'(0)$, où φ_h est définie comme en 4.3 2.

Définition 4.14. Soit $f : U \to \mathbb{R}^m$, si les dérivées partielles de f existent en tout point de U et sont continues de U dans $\mathbb{R}^m$, on dit que f est *de classe* $\mathcal{C}^1$ sur U.

Exemple 4.15. Les exemples de 4.9 sont de classe $\mathcal{C}^1$ sur leurs ensembles de définition.

Théorème 4.16. *Si f est de classe $\mathcal{C}^1$ sur U, alors f est différentiable en tout point de U.*

Remarque 4.17. Ce théorème est faux, si on demande seulement que les dérivées partielles soient continues en chacune des variables indépendamment. Un contre-exemple est donné dans le TD, fiche 2 exercice 5.

4.3 Opérations sur les fonctions vectorielles différentiables

Proposition 4.18. *Soient f et $\tilde{f} : U \to \mathbb{R}^m$, $a \in U$ et $\lambda \in \mathbb{R}$. Si f et $\tilde{f}$ sont différentiables en a (resp. sont de classe $\mathcal{C}^1$ sur U), alors $f + \tilde{f}$, et λf sont différentiables en a (resp. sont de classe $\mathcal{C}^1$ sur U).*
Dans ce cas, en $x = a$ (resp. pour tout $x \in U$), on a :

$$[d(f+\tilde{f})](x) = [df](x) + [d\tilde{f}](x) \quad \textit{et} \quad [d(\lambda f)](x) = \lambda[df](x).$$

Proposition 4.19. *Soient $f : U \to \mathbb{R}^m$, $\alpha : U \to \mathbb{R}$ et $a \in U$. Si f et α sont différentiables en a (resp. sont de classe $\mathcal{C}^1$ sur U), alors αf est différentiable en a (resp. de classe $\mathcal{C}^1$ sur U).*
Dans ce cas, en $x = a$ (resp. pour tout $x \in U$), on a :

$$\forall h \in \mathbb{R}^n, \quad [d(\alpha f)](x).h = ([d\alpha](x).h)f(x) + \alpha(x)([df](x).h).$$

Remarque 4.20. Les formules de 4.18 sont aussi valables pour les jacobiennes et les dérivées partielles. Celle de 4.19 est valable pour les dérivées partielles. La traduire en termes matriciels demande du soin.

Proposition 4.21 (Chain Rule)**.** *Soient $f : U \to V$ et $g : V \to \mathbb{R}^p$ avec $V \subset \mathbb{R}^m$. Si f est différentiable en $a \in U$ (resp. $\mathcal{C}^1$ sur U) et g est différentiable en $f(a) \in V$ (resp. $\mathcal{C}^1$ sur V), alors $g \circ f$ est différentiable en a (resp. $\mathcal{C}^1$ sur U). Dans ce cas, pour $x = a$ (resp. pour tout $x \in U$) on a :*

$$[d(g \circ f)](x) = [dg](f(x)) \circ [df](x).$$

Remarque 4.22. Ceci se traduit sur les jacobiennes par : $J(g \circ f)(x) = [J(g)(f(x))][J(f)(x)]$.

Utilisons cette relation en un point a vérifiant les hypothèses de 4.21. En notant $(y_1, \ldots, y_m)$ les coordonnées dans l'espace intermédiaire $\mathbb{R}^m$ contenant V, on obtient, pour tout $1 \leqslant i \leqslant p$ et tout $1 \leqslant j \leqslant n$, la relation :

$$\frac{\partial (g \circ f)_i}{\partial x_j}(a) = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(f(a)) \frac{\partial f_k}{\partial x_j}(a).$$

Calcul des dérivées partielle d'une composée. Pour calculer $\frac{\partial (g\circ f)_i}{\partial x_j}(a)$, on utilise la formule ci-dessus, après avoir calculer les dérivées partielles de f et g qui y apparaissent (exemple en TD, fiche 3 exercice 3).

Calcul de la différentielle d'une composée. On calcule les dérivées partielles de f en a et de g en $f(a)$, ce qui fournit leurs jacobiennes. On calcule ensuite $J(g \circ f)(a)$ par 4.22, ce qui est équivalent à calculer la différentielle. On peut aussi utiliser $[d(g \circ f)](a) = \sum_{j=1}^{n} \frac{\partial (g\circ f)}{\partial x_j}(a)dx_j$, et utiliser la formule ci-dessus, ce qui revient au même.

Classe $\mathcal{C}^1$. Souvent, pour montrer que $f : U \to \mathbb{R}^m$ est de classe $\mathcal{C}^1$, on remarque qu'elle est obtenue (ou que chacune de ses composantes est obtenue) par opérations élémentaires (sommes, produits, compositions) à partir de fonctions vectorielles de classe $\mathcal{C}^1$ (par exemple celles de 4.9) et est donc de classe $\mathcal{C}^1$ par les propositions 4.18, 4.19 et 4.21. On peut aussi calculer ses dérivées partielles et montrer qu'elles sont continues sur U.

Exemple 4.23. La fonction $x \mapsto \|x\|^2 = \sum (x_j)^2$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$. Sa racine carrée : $x \mapsto \|x\|$, est de classe $\mathcal{C}^1$ sur $\mathbb{R}^n \setminus \{0\}$. Cependant, elle n'est pas différentiable en 0 (où on a noté 0 pour $(0, \ldots, 0)$).

Non différentiabilité en un point. Soit $f : U \to \mathbb{R}^m$ et $a \in U$. On commence par vérifier si f est continue en a. Si elle ne l'est pas, elle n'y est pas différentiable d'après 4.10 (exemple en 4.13).

On vérifie que f est dérivable dans toute direction en a. Si f n'est pas dérivable dans la direction h en a, elle ne peut être différentiable d'après 4.12.

Si f est dérivable dans toutes les directions en a, la seule application qui pourrait être la différentielle de f en a est $h \mapsto [D_h f](a)$, d'après 4.12. Si c'était effectivement la différentielle, elle serait linéaire et on aurait : $\forall h \in \mathbb{R}^n$, $[D_h f](a) = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a)$. Il suffit alors de trouver $h \in \mathbb{R}^n$ pour lequel cette formule n'est pas satisfaite pour conclure que f n'est pas différentiable en a (exemple en TD, fiche 2 exercice 5.4).

Différentiabilité en un point. Si f est de classe $\mathcal{C}^1$, on peut conclure par le théorème 4.16. On peut aussi remarquer que f s'obtient par opérations élémentaires à partir de fonctions vectorielles différentiables en a et conclure par 4.18, 4.19 et 4.21. Si cela échoue, on commence par déterminer le seul candidat pour la différentielle, $L : h \mapsto \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a)$. Il s'agit ensuite de vérifier qu'on a bien $\frac{f(a+h)-f(a)-L(h)}{\|h\|} \xrightarrow[h\to 0]{} 0$.

4.4 Gradient

Dans le cas particulier où $m = 1$, soit $f : U \to \mathbb{R}$ différentiable en $a \in \mathbb{R}^n$. On a $[df](a) : \mathbb{R}^n \to \mathbb{R}$ linéaire, donc d'après la section 3, il existe un unique $b \in \mathbb{R}^n$ tel que $[df](a) : h \mapsto b \cdot h = \sum_{j=1}^{n} b_j h_j$.

Définition 4.24. Soit $f : U \to \mathbb{R}$ différentiable en $a \in U$, on appelle *gradient* de f en a, et on note $\nabla f(a)$ (ou $\texttt{grad}(f)(a)$) l'unique vecteur de $\mathbb{R}^n$ tel que : $\forall h \in \mathbb{R}^n,\ [df](a).h = (\nabla f(a)) \cdot h$.

Calcul du gradient. On a $\nabla f(a) = (\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) e_j$, où les $\frac{\partial f}{\partial x_j}(a)$ sont réels car $f : U \to \mathbb{R}$. On calcule donc facilement le gradient à partir des dérivées partielles.

Remarques 4.25.
1. La formule ci-dessus a du sens dès que f a des dérivées partielles en a, mais la notion de gradient n'a de sens que lorsque f est différentiable en a, et est à valeurs dans $\mathbb{R}$.
2. Géométriquement, si $\nabla f(a) \neq 0$, c'est la direction de plus forte croissance de f en a. Il est orthogonal en a à la ligne de niveau $f(a)$: $L_{f(a)} = \{x \in \mathbb{R}^n | f(x) = f(a)\}$.

4.5 Cas où $n = 1$

Lorsque $a \in U \subset \mathbb{R}$ et $f : U \to \mathbb{R}^m$, plusieurs objets coexistent : la dérivée $f'(a) = (f_1'(a), \ldots, f_m'(a)) \in \mathbb{R}^m$, la différentielle $[df](a) : \mathbb{R} \to \mathbb{R}^m$, et l'unique dérivée partielle $\frac{\partial f}{\partial x}(a) \in \mathbb{R}^m$ (sous réserve d'existence).

Proposition 4.26. *Soient $f : U \to \mathbb{R}^m$ avec $U \subset \mathbb{R}$ et $a \in U$, les conditions suivantes sont équivalentes :*
1. *f est dérivable en a,*
2. *f a une dérivée partielle en a (selon l'unique variable x),*
3. *f est différentiable en a.*

Lorsque ces conditions sont vérifiées, on a $f'(a) = \frac{\partial f}{\partial x}(a)$ et $[df](a) : h \mapsto hf'(a)$.

Cas de la dérivée d'une composée. Soient $U \subset \mathbb{R}$, $V \subset \mathbb{R}^m$, $f : U \to V$, $g : V \to \mathbb{R}^p$ et $a \in U$. Si f est dérivable en a et g est différentiable en $f(a)$, alors $(g \circ f)$ est dérivable en a d'après 4.26 et 4.21. De plus (voir par exemple TD 3 exercices 4 et 6), on a l'égalité suivante dans $\mathbb{R}^p$:

$$(g \circ f)'(a) = [dg](f(a)).(f'(a)) = \sum_{k=1}^{m} \frac{\partial g}{\partial x_k}(f(a)) f_k'(a).$$

5 Fonctions de classe $\mathcal{C}^2$

Définition 5.1. Soit $f : U \to \mathbb{R}$, on dit que f est de *classe $\mathcal{C}^2$* sur U, si ses dérivées partielles existent et sont de classe $\mathcal{C}^1$ sur U. On appelle alors *dérivées (partielles) secondes* les dérivées partielles des fonctions $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$, et on note pour tout i et$j \in \{1, \ldots, n\}$, pour tout $x \in U$:

$$\frac{\partial^2 f}{\partial x_i^2}(x) = \frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_i})(x) \quad \text{et} \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})(x).$$

Classe $\mathcal{C}^2$. Pour prouver qu'une fonction est $\mathcal{C}^2$, on a des résultats similaires à 4.18, 4.19 et 4.21 dans le cas $\mathcal{C}^2$. Une fonction obtenue par opérations élémentaires à partir de fonctions $\mathcal{C}^2$ est donc $\mathcal{C}^2$. On utilise ceci et le fait que les fonctions des exemples 1.13 et 4.9 sont de classe $\mathcal{C}^2$ sur leur domaine de définition.

Théorème 5.2 (Schwarz)**.** *Si f est $\mathcal{C}^2$, alors pour tout i et $j \in \{1,\ldots,n\}$, $\forall x \in U$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$.*

Remarques 5.3.
1. Pour calculer les dérivées secondes de f en a, il faut connaître ses dérivées partielles sur un voisinage de a, et pas seulement au point a.
2. On peut définir les dérivées secondes de f en a sous des hypothèses plus faibles que "f de classe $\mathcal{C}^2$", mais le théorème 5.2 n'est plus vrai.

Définition 5.4. Si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^2$, on appelle la matrice $\mathtt{Hess}(f)(a) \in \mathcal{M}_{n,n}$ dont les coefficients sont les $(\frac{\partial^2 f}{\partial x_i \partial x_j}(a))$ la *hessienne* de f en a.

Calcul des dérivées secondes et de la hessienne. Pour calculer $\frac{\partial^2 f}{\partial x_i \partial x_j}(a)$, on dérive successivement f par rapport à x_j et x_i en traitant à chaque fois les autres variables comme des constantes. Pour calculer la hessienne $\mathtt{Hess}(f)(a)$, on calcule toutes les dérivées secondes de f en a. D'après le théorème 5.2, il suffit de le faire pour les indices $i \leqslant j$. On obtient les autres termes par symétrie.

Proposition 5.5 (Formule de Taylor à l'ordre 1)**.** *Soit $f : U \to \mathbb{R}$ différentiable en $a \in U$, on a :*

$$f(a+h) = f(a) + \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a) + \underset{h\to 0}{o}(\|h\|), \qquad \textit{ou encore,} \tag{5.1}$$

$$f(x) = f(a) + \sum_{j=1}^{n} (x_j - a_j) \frac{\partial f}{\partial x_j}(a) + \underset{x\to a}{o}(\|x-a\|). \tag{5.2}$$

Proposition 5.6 (Formule de Taylor à l'ordre 2)**.** *Soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in U$, on a :*

$$f(a+h) = f(a) + \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a) + \frac{1}{2} \sum_{\substack{1\leqslant i\leqslant n \\ 1\leqslant j\leqslant n}} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + \underset{h\to 0}{o}(\|h\|^2), \qquad \textit{ou encore,} \tag{5.3}$$

$$f(x) = f(a) + \sum_{j=1}^{n} (x_j - a_j) \frac{\partial f}{\partial x_j}(a) + \frac{1}{2} \sum_{\substack{1\leqslant i\leqslant n \\ 1\leqslant j\leqslant n}} (x_i - a_i)(x_j - a_j) \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + \underset{h\to 0}{o}(\|x-a\|^2). \tag{5.4}$$

Développement de Taylor. On appelle *partie principale du développement de Taylor à l'ordre $q \in \{1,2\}$ de f en a* les termes précédant le petit o dans les parties droites des formules ci-dessus. Pour les calculer, on calcule les dérivées partielles qui apparaissent dans la formule, donc jusqu'à l'ordre q.

Remarques 5.7.
1. La partie principale du développement de Taylor de f en a à l'ordre q est un polynôme de degré q en n variables ($(h_1,\ldots,h_n)$ pour la première forme, $(x_1,\ldots,x_n)$ pour la seconde). C'est le polynôme de degré au plus q en n variables qui approche "le mieux" f au voisinage de a.
2. En particulier, si f est un polynôme de degré inférieur à q, il est égal à la partie principale de son développement de Taylor à l'ordre q en un point quelconque.
3. On dit que f est *de classe $\mathcal{C}^k$ sur U*, si toutes ses dérivées partielles jusqu'à l'ordre k existent et sont continues sur U. L'ordre dans lequel on effectue les dérivations n'a pas d'importance d'après 5.2. On a alors des formules de Taylor pour f à tout ordre $q \in \{1,\ldots,k\}$ et en tout point $a \in U$. On dit que f est *de classe $\mathcal{C}^\infty$ sur U* si elle y est de classe $\mathcal{C}^k$ pour tout $k \in \mathbb{N}$.
4. Le contenu de cette section s'étend au cas des fonctions vectorielles $f : U \to \mathbb{R}^m$, sauf la notion de hessienne.

6 Points critiques, extrema

Définition 6.1. Soit $f : U \to \mathbb{R}$ et $a \in U$, on dit que a est un *maximum* (resp. un *minimum*) *global* de f si $\forall x \in U$, $f(x) \leqslant f(a)$ (resp. $f(x) \geqslant f(a)$).
On dit que a est un *maximum* (resp. un *minimum*) *local* de f s'il existe $r > 0$ tel que $B(a,r) \subset U$ et $\forall x \in B(a,r)$, $f(x) \leqslant f(a)$ (resp. $f(x) \geqslant f(a)$).
On dit que a est un *extremum global* (resp. *local*) de f, si a est un maximum global (resp. local) ou un minimum global (resp. local) de f.

Définition 6.2. Soit $f : U \to \mathbb{R}^m$ différentiable $a \in U$, on dit que a est un *point critique* de f lorsque $[df](a)$ est l'application linéaire nulle, c'est-à-dire si toutes les dérivées partielles de f s'annulent en a.

Proposition 6.3. *Soit $f : U \to \mathbb{R}$ différentiable $a \in U$, si a est un extremum local de f alors a est un point critique de f, c'est-à-dire $[df](a)$ est nulle, ou encore $\nabla f(a) = 0$.*

Remarques 6.4.
1. La réciproque est fausse, par exemple 0 est critique pour $x \mapsto x^3$ de $\mathbb{R}$ dans $\mathbb{R}$.
2. Un extremum global est en particulier en extremum local.
3. La notion d'extremum n'a de sens que pour les fonctions, alors que celle de point critique a du sens pour toutes les fonctions vectorielles.

Détermination des points critiques. Pour déterminer les points critiques d'une fonction vectorielle différentiable $f : U \to \mathbb{R}^m$, on commence par calculer ses dérivées partielles, puis on cherche les points $x \in \mathbb{R}^n$ où elles s'annulent simultanément. On a généralement à résoudre un système compliqué d'équations.

Détermination des extrema. On n'a pas de méthode générale pour déterminer les extrema d'une fonction $f : U \to \mathbb{R}$ qui n'est pas au moins différentiable. Si f est différentiable, on calcule ses points critiques, ceux où ∇f s'annule, puis on détermine lesquels sont des extrema locaux. Il faut ensuite regarder les valeurs de f sur les différents extrema locaux pour déterminer si certains sont globaux.

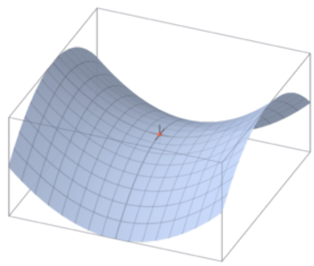
Remarque 6.5. Une fonction peut ne pas avoir d'extrema. Elle peut aussi avoir des extrema locaux mais pas d'extrema globaux, par exemple $f : \mathbb{R} \to \mathbb{R}$ définie par $x \mapsto x^3 - x$.

Définition 6.6. Soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in U$ un point critique de f. On dit que a est un point critique *dégénéré* si $\det(\mathtt{Hess}(f)(a)) = 0$ et *non dégénéré* sinon.

Définition 6.7. Un point critique non dégénéré qui n'est pas un extremum local est appelé un *point selle* ou *point col.*

Soient $U \subset \mathbb{R}^2$, $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in U$ un point critique non dégénéré de f.

Proposition 6.8. *Si* $\det(\mathtt{Hess}(f)(a)) > 0$*, le point a est un extremum local de f, si* $\det(\mathtt{Hess}(f)(a)) < 0$*, le point a est un point col de f.*



Graphe d'une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ au voisinage d'un point col.
Source : Wikipedia.

Comme $\det(\mathtt{Hess}(f)(a)) = \frac{\partial^2 f}{\partial x^2}(a)\frac{\partial^2 f}{\partial y^2}(a) - (\frac{\partial^2 f}{\partial x \partial y}(a))^2$ (voir section 3), si cette quantité est positive, alors $\frac{\partial^2 f}{\partial x^2}(a)$ et $\frac{\partial^2 f}{\partial y^2}(a)$ sont non nuls et de même signe.

Proposition 6.9. *Si $\frac{\partial^2 f}{\partial x^2}(a)$ et $\frac{\partial^2 f}{\partial y^2}(a)$ sont positifs alors a est un minimum local, s'ils sont négatifs, a est un maximum local.*

Nature des points critiques. Dans le cas d'une fonction $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$ avec $U \subset \mathbb{R}^2$ et $a \in U$ un point critique de f non dégénéré, on peut déterminer si a est un maximum local, un minimum local ou un point selle. On calcule les dérivées secondes de f en a, puis $\det(\mathtt{Hess}(f)(a))$, et on applique 6.8 et 6.9.

Remarque 6.10. Lorsque a est un point critique dégénéré on ne peut rien dire. Il existe des exemples avec a minimum local, avec a maximum local et avec a qui n'est pas un extremum local.

Définition 6.11. Soient $g : U \to \mathbb{R}$ de classe $\mathcal{C}^1$ et $z \in \mathbb{R}$, on note $L_z = \{x \in U | g(x) = z\}$ la ligne de niveau z. On dit que z est un niveau régulier de g si pour tout $x \in L_z$, $\nabla g(x) \neq 0$.

Théorème 6.12 (des extrema liés). *Soient f et $g : U \to \mathbb{R}$ de classe $\mathcal{C}^1$ et $z \in \mathbb{R}$ un niveau régulier de g. Si la restriction de f à L_z atteint un extremum local en $x \in L_z$, alors il existe $\lambda \in \mathbb{R}$ tel que $\nabla f(x) = \lambda \nabla g(x)$. Le réel λ est appelé* multiplicateur de Lagrange*.*

Remarque 6.13. Les points $x \in L_z$ tels qu'il existe $\lambda \in \mathbb{R}$ vérifiant $\nabla f(x) = \lambda \nabla g(x)$ sont les points critiques de f restreinte à L_z. Ce ne sont pas tous des extrema de la restriction.

Recherche d'extrema liés. On utilise ce théorème pour déterminer les points critiques d'une fonction de classe $\mathcal{C}^1$, $f : \mathbb{R}^2 \to \mathbb{R}$, restreinte à une courbe (resp. $f : \mathbb{R}^3 \to \mathbb{R}$ restreinte à une surface) qui est définie par une équation du type $g(x) = z$ où z est une constante et g est $\mathcal{C}^1$. Pour cela, on calcule les gradients de f et de g en tout point de la courbe (resp. de la surface) et on détermine en quels points il existe un multiplicateur de Lagrange. Il reste ensuite à déterminer, lesquels parmi ces points sont des extrema locaux, globaux. Un exemple est donné dans le dernier exercice de la fiche 4 du TD.