QUELQUES PREUVES DEMANDÉES EN DM ET AUTRES

Comme dans d'autres rédactions, je donnerai pour un certain nombre de preuves deux versions : d'abord une version longue pour expliquer le raisonnement en tout détail, suivi d'une version "courte" qui suffit dans une rédaction officielle.

Preuve de 1.14. En principe cette preuve est "trop trivial" pour être incluse, mais je la rajoute pour montrer comment manipuler des va-et-vient entre $\varepsilon > 0$ (pour une boule $B_\varepsilon(x)$) et un voisinage de x.

L'équivalence se montre par double implication et on commence avec l'implication "x adhérent à $A \Rightarrow$ pour tout voisinage ouvert de x on a $U \cap A \neq \emptyset$." On a donc le schéma "on sait — on veut montrer"

$$\begin{array}{rl} \text{on sait} & \forall \varepsilon > 0 : B_\varepsilon(x) \cap A \neq \emptyset \\ \text{on veut montrer} & \forall U \text{ voisinage ouvert de } x : U \cap A \neq \emptyset \ . \end{array}$$

Pour le montrer on prend U un voisinage ouvert de x arbitraire et on cherche à montrer qu'on a $U \cap A \neq \emptyset$. En utilisant la définition d'un ouvert, on a $x \in U$ et U ouvert, donc il existe $r > 0$ tel que $B_r(x) \subset U$. Et on sait quelque chose pour tout $\varepsilon > 0$, donc en particulier pour $\varepsilon = r > 0$, à savoir $B_r(x) \cap A \neq \emptyset$. On a donc

$$\emptyset \neq B_r(x) \cap A \subset U \cap U \ ,$$

ce qui montre que $U \cap A \neq \emptyset$ comme souhaité.

Réciproquement, on a le schéma "on sait — on veut montrer"

$$\begin{array}{rl} \text{on sait} & \forall U \text{ voisinage ouvert de } x : U \cap A \neq \emptyset \\ \text{on veut montrer} & \forall \varepsilon > 0 : B_\varepsilon(x) \cap A \neq \emptyset \ . \end{array}$$

Pour le montrer on prend donc $\varepsilon > 0$ arbitraire. Alors $B_\varepsilon(x)$ sera un ouvert qui contient x ; c'est donc un voisinage ouvert de x. Et on sait quelque chose pour tout voisinage ouvert U de x, donc en particulier pour $U = B_\varepsilon(x)$. Par ce qu'on sait, on en déduit qu'on a $B_\varepsilon(x) \cap A \neq \emptyset$.

Preuve de 2.5. J'ai donné cette preuve en cours, mais je le répète ici. Et comme pour beaucoup d'autres preuves, on démontre l'équivalence par une double implication, en commençant avec l'implication $\lim_{n\to\infty} x_n = \ell \ \Rightarrow\ \forall i : \lim_{n\to\infty} x_n^i = \ell^i$. On a donc le schéma "on sait — on veut montrer"

$$\begin{array}{rl} \text{on sait} & \forall \varepsilon > 0 \ \exists N \in \mathbf{N} \ \forall n \geq N : \|x_n - \ell\| < \varepsilon \\ \text{on veut montrer} & \forall i \ \forall \varepsilon' > 0 \ \exists N' \in \mathbf{N} \ \forall n' \geq N' : \left|x_{n'}^i - \ell^i\right| < \varepsilon' \ . \end{array}$$

Pour montrer ce qu'on veut montrer, on prend "donc" i (entre 1 et p) et $\varepsilon' > 0$ arbitraires et on cherche $N' \in \mathbf{N}$ avec la propriété

$$\forall n' \geq N' : \left|x_{n'}^i - \ell^i\right| < \varepsilon' \ .$$

Pour le trouver, il faut “évidemment” se servir de ce qu’on sait. Il faut “donc” relier (si possible) le résultat $\left|x^i_{n'} - \ell^i\right| < \varepsilon'$ au résultat de ce qu’on sait, à savoir $\|x_n - \ell\| < \varepsilon$. Mais comment ? C’est à ce moment qu’on réalise qu’on ne sait pas quelle norme on est censé d’utiliser. Mais ... on peut utiliser la norme qu’on veut ! À condition de faire le choix au tout début d’un raisonnement et de ne plus le changer en cours de route. Comme j’ai expliqué souvent, on **pense** le plus souvent en termes de la norme $\|\cdot\|_2$, la norme euclidienne, mais pour les calculs, le choix de la norme $\|\cdot\|_1$ ou $\|\cdot\|_\infty$ est plus souvent beaucoup plus facile à utiliser dans les calculs. D’autre part, pour un certain nombre de calculs ce choix ne change quasiment rien. Et c’est le cas ici. Car pour les trois normes on a l’inégalité, pour tout $i = 1, \ldots, p$,

$$x = (x^1, \ldots, x^p) \in \mathbf{R}^p \quad : \quad \begin{cases} |x^i| \le \displaystyle\sum_{j=1}^p |x^j| = \|x\|_1 \\ |x^i| = \sqrt{(x^i)^2} \le \sqrt{\displaystyle\sum_{j=1}^p (x^j)^2} = \|x\|_2 \\ |x^i| \le \displaystyle\max_{j=1,\ldots,p} |x^j| = \|x\|_\infty \ . \end{cases}$$

Pour chacune des trois possibilités on a donc l’inégalité $\left|x^i_{n'} - \ell^i\right| \le \|x_n - \ell\|$. Avec cela en tête, on rappelle qu’on sait quelque chose pour tout $\varepsilon > 0$, donc en particulier pour le choix $\varepsilon = \varepsilon'$. Pour ce choix de ε il existe donc (par ce qu’on sait) $N \in \mathbf{N}$ avec la propriété

$$\forall n \ge N : \|x_n - \ell\| < \varepsilon \ .$$

En voyant cela, on “voit” que le choix $N' = N$ convient, car pour $n \ge N' = N$ on peut faire le raisonnement

$$\left|x^i_{n'} - \ell^i\right| \le \|x_n - \ell\| \le \|x_n - \ell\| < \varepsilon = \varepsilon' \ .$$

Réciproquement, on a le schéma “on sait — on veut montrer”

$$\begin{array}{rl} \text{on sait} & \forall i\ \forall \varepsilon' > 0\ \exists N' \in \mathbf{N}\ \forall n' \ge N' : \left|x^i_{n'} - \ell^i\right| < \varepsilon' \\ \text{on veut montrer} & \forall \varepsilon > 0\ \exists N \in \mathbf{N}\ \forall n \ge N : \|x_n - \ell\| < \varepsilon \ . \end{array}$$

Pour montrer ce qu’on veut montrer, on prend “donc” $\varepsilon > 0$ arbitraire et on cherche $N \in \mathbf{N}$ avec la propriété

$$\forall n \ge N : \|x_n - \ell\| < \varepsilon \ .$$

Et on est de nouveau confronté avec la question quelle norme utiliser. Contrairement à l’implication précédente, ici le choix de la norme euclidienne nous désavantage. On choisit “donc” la norme $\|\cdot\|_\infty$ (le choix de $\|\cdot\|_1$ donnera un raisonnement assez analogue et pas plus difficile, mais il faut faire un choix !). On cherche donc $N \in \mathbf{N}$ avec la propriété

$$\forall n \ge N : \|x_n - \ell\| \equiv \max_{j=1,\ldots,p} \left|x^j_{n'} - \ell^j\right| < \varepsilon \ .$$

Et justement, dans ce qu’on sait on trouve quelque chose sur $\left|x^i_{n'} - \ell^i\right| < \varepsilon'$. On va donc dire qu’on sait quelque chose pour tout $\varepsilon' > 0$ et donc en particulier pour la valeur $\varepsilon' = \varepsilon$. **Mais** ... dans ce qu’on sait il y a aussi un “pour tout i,” qu’on ne peut pas oublier ! Il faut donc aussi préciser quelle valeur pour i on prend ! Sachant qu’on doit calculer $\max_j$, il semble indiqué d’utiliser une valeur “arbitraire.” Pour nos choix de i et de $\varepsilon > 0$ on trouve donc N' avec la propriété

$$\forall n' \ge N' : \left|x^i_{n'} - \ell^i\right| < \varepsilon' \ .$$

Mais attention : ce N' on l'obtient **après** avoir fait un choix pour i ! Ce qui veut dire que le N' qu'on obtient peut être différent quand on prend un autre i. Autrement dit, ce N' qu'on trouve dépend de la valeur de i ! On a donc intérêt de la noter comme N'_i. En conclusion : pour $\varepsilon' = \varepsilon$ et i arbitraire on trouve N'_i avec la propriété

$$\forall n' \geq N'_i : \left| x^i_{n'} - \ell^i \right| < \varepsilon' \ .$$

Rappelons maintenant qu'on cherche N avec la propriété

$$\forall n \geq N : \|x_n - \ell\| \equiv \max_{j=1,\dots,p} \left| x^j_{n'} - \ell^j \right| < \varepsilon \ .$$

Si on veut être sûr que tous les $\left| x^j_{n'} - \ell^j \right|$ sont majorés par $\varepsilon = \varepsilon'$, il faut "donc" choisir n plus grand (ou égal) que tous les N'_i. Ceci nous amène à poser/choisir

$$N = \max_{i=1,\dots,p} N'_i \ ,$$

et de faire le raisonnement

$$n \geq N \ \Rightarrow\ \forall j : n \geq N'_j \ \Rightarrow\ \forall j : \left| x^j_{n'} - \ell^j \right| < \varepsilon' = \varepsilon \ \Rightarrow\ \max_{j=1,\dots,p} \left| x^j_{n'} - \ell^j \right| < \varepsilon' = \varepsilon \ ,$$

ce qu'on résumer dans l'implication

$$n \geq N \ \Rightarrow\ \|x\|_\infty = \max_{j=1,\dots,p} \left| x^j_{n'} - \ell^j \right| < \varepsilon' = \varepsilon \ ,$$

ce qui est la conclusion souhaitée.

Rédaction "courte" officielle. On choisit d'utiliser (partout) la norme $\|\cdot\|_\infty$ sur $\mathbf{R}^p$. Pour l'implication directe, prenons i entre 1 et p et $\varepsilon > 0$. Alors par définition de $\lim_{n\to\infty} x_n = \ell$ il existe $N \in \mathbf{N}$ avec la propriété

$$\forall n \geq N : \|x_n - \ell\| < \varepsilon \ .$$

On a donc, avec notre i, le raisonnement

$$n \geq N \ \Rightarrow\ \left| x^i_{n'} - \ell^i \right| \leq \|x_n - \ell\|_\infty < \varepsilon \ ,$$

ce qui montre qu'on a bien $\lim_{n\to\infty} x^i_n = \ell^i$.

Réciproquement, soit $\varepsilon > 0$ arbitraire. Alors par définition de $\lim_{n\to\infty} x^i_n = \ell^i$ il existe pour tout i un $N_i \in \mathbf{N}$ avec la propriété

$$\forall n \geq N_i : \left| x^i_{n'} - \ell^i \right| < \varepsilon \ .$$

On pose maintenant $N = \max_{i=1,\dots,p} N_i$ et on calcule :

$$n \geq N \ \Rightarrow\ \forall i : n \geq N'_j \ \Rightarrow\ \forall i : \left| x^j_{n'} - \ell^j \right| < \varepsilon' = \varepsilon \ \Rightarrow\ \max_{i=1,\dots,p} \left| x^i_{n'} - \ell^i \right| < \varepsilon \ ,$$

ce qui est la définition de $\lim_{n\to\infty} x_n = \ell$ comme souhaité.

Preuve de 2.8. Cette preuve est "complètement trivial," mais il faut savoir pourquoi c'est trivial, car il y a quand même quelque chose à montrer. La différence entre les deux propriétés est que la première parle d'une limite dans l'espace vectoriel F, tandis que la deuxième parle d'une limite dans l'espace vectoriel $\mathbf{R}$. Et dans $\mathbf{R}$ on utilise la valeur absolue comme norme. Quand on écrit rigoureusement cette dernière propriété, on établit directement les équivalences (parce que une norme est positive !) :

$$\begin{aligned} &\lim_{x\to a} \|f(x) - \ell\| = 0 \\ &\overset{\text{déf.}}{\iff} \quad \forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in A : 0 < \|x - a\| < \delta \ \Rightarrow\ \big|\, \|f(x) - \ell\| - 0 \,\big| < \varepsilon \end{aligned}$$

$$\begin{aligned} &\iff \forall\varepsilon > 0\ \exists\delta > 0\ \forall x \in A : 0 < \|x-a\| < \delta \ \Rightarrow\ \|f(x) - \ell\| < \varepsilon \\ &\overset{\text{déf.}}{\iff} \lim_{x\to a} f(x) = \ell \ . \end{aligned}$$

Preuve de 6.1. Cette preuve ne demande aucune astuce particulière, simplement un calcul direct. La définition de la norme infinie d'une matrice est comme pour un vecteur : la plus grande valeur parmi les valeurs absolues des éléments de matrices A_{ij} :

$$\|A\|_\infty = \max_{i=1,\dots p;j=1;\dots,n} |A_{ij}| \ .$$

Quand on écrit $y = Ax$, ceci veut dire que les composantes sont données par

$$y_i = \sum_{j=1}^n A_{ij}\, x_j \ .$$

On a donc directement

$$\begin{aligned} \|Ax\|_1 = \|y\|_1 = \sum_{i=1}^p |y_i| = \sum_{i=1}^p \left|\sum_{j=1}^n A_{ij}\, x_j\right| &\le \sum_{i=1}^p\sum_{j=1}^n |A_{ij}|\cdot|x_j| \\ &\le \sum_{i=1}^p\sum_{j=1}^n \|A\|_\infty\cdot|x_j| = \|A\|_\infty\cdot\sum_{i=1}^p\sum_{j=1}^n |x_j| = \|A\|_\infty\cdot\sum_{i=1}^p \|x\|_1 \\ &= p\cdot\|A\|_\infty\cdot\|x\|_1 \ , \end{aligned}$$

ainsi que

$$\begin{aligned} \|Ax\|_\infty = \|y\|_\infty = \max_{i=1,\dots,p} |y_i| = \max_{i=1,\dots,p} \left|\sum_{j=1}^n A_{ij}\, x_j\right| &\le \max_{i=1,\dots,p} \sum_{j=1}^n |A_{ij}|\cdot|x_j| \\ &\le \max_{i=1,\dots,p} \sum_{j=1}^n \|A\|_\infty\cdot\|x\|_\infty = \|A\|_\infty\cdot\|x\|_\infty \max_{i=1,\dots,p}\sum_{j=1}^n 1 \\ &= n\cdot\|A\|_\infty\cdot\|x\|_\infty \ . \end{aligned}$$

Preuve de 6.4. Appliquons encore une fois le schéma "on sait — on veut montrer"

$$\begin{aligned} &\text{on sait} && \lim_{x\to a}\frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|}{\|x-a\|} = 0 \\ &\text{on veut montrer} && \forall\varepsilon > 0\ \exists\delta > 0\ \forall x \in U : \|x-a\| < \delta \ \Rightarrow\ \|f(x)-f(a)\| < \varepsilon \ . \end{aligned}$$

Pour "rapprocher" les deux propriétés, on remarque que dans un ouvert $U \subset \mathbf{R}^n$ tout point $a \in U$ est un point d'accumulation de U, ce qui permet d'appliquer 2.16 et d'affirmer que la propriété à montrer est équivalente à la propriété

$$\text{on veut montrer} \qquad \lim_{x\to a} f(x) = f(a) \ .$$

Et on peut aller encore plus loin et appliquer 2.1 pour affirmer que c'est équivalente à la propriété

$$\text{on veut montrer} \qquad \lim_{x\to a} \|f(x) - f(a)\| = 0 \ .$$

Écrit comme cela, les deux propriétés (on sait — on veut montrer) se ressemblent beaucoup et il "suffit" maintenant de faire le lien entre les deux. Autrement dit, on cherche à exprimer $f(x)-f(a)$ en termes de $\big(f(x)-f(a)-\big((Df)(a)\big)(x-a)\big)/\|x-a\|$. Ce qui n'est pas très difficile :

$$\begin{aligned} f(x)-f(a) &= f(x)-f(a)-\big((Df)(a)\big)(x-a)+\big((Df)(a)\big)(x-a) \\ &= \frac{f(x)-f(a)-\big((Df)(a)\big)(x-a)}{\|x-a\|}\cdot\|x-a\|+\big((Df)(a)\big)(x-a)\ , \end{aligned}$$

ce qui nous donne la majoration

$$\begin{aligned} \|f(x)-f(a)\| &= \left\|\frac{f(x)-f(a)-\big((Df)(a)\big)(x-a)}{\|x-a\|}\cdot\|x-a\|+\big((Df)(a)\big)(x-a)\right\| \\ &\leq \frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|}{\|x-a\|}\cdot\|x-a\| \\ &\qquad +\Big\|\big((Df)(a)\big)(x-a)\Big\|\ . \end{aligned} \tag{0.1}$$

Quand on regarde ce résultat, on peut invoquer le théorème des gendarmes : on a

$$\lim_{x\to a}\frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|}{\|x-a\|}=0 \qquad \text{et} \qquad \lim_{x\to a}\|x-a\|=0$$

$$\text{donc} \qquad \lim_{x\to a}\frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|}{\|x-a\|}\cdot\|x-a\|=0\cdot 0=0\ .$$

Mais on sait aussi que $(Df)(a)$ est une application linéaire/matrice, ce qui est un application continue (en dimension finie !), donc par continuité de $(Df)(a)$ et le fait que prendre la norme est aussi une opération continue 2.23, on peut appliquer 2.27 sur la composée de deux applications continues et conclure qu'on a

$$\lim_{x\to a}\Big\|\big((Df)(a)\big)(x-a)\Big\|=\|0\|=0\ .$$

Le membre de droite dans (0.1) tend donc vers 0 dans la limite $x\to a$. Sachant que $\|f(x)-f(a)\|$ est toujours positive, le théorème de gendarmes s'applique et on peut conclure qu'on a aussi

$$\lim_{x\to a}\|f(x)-f(a)\|=0\ ,$$

ce qu'on voulait montrer.

Au lieu d'invoquer la continuité de l'application linéaire $(Df)(a)$, on aurait pu utiliser 6.1. Pour cela il suffit de commencer avec la remarque qu'on peut faire les calculs avec une norme de notre choix. Choisissons la norme $\|\cdot\|_\infty$ (dès le début !) et appliquons 6.1 qui donne :

$$\Big\|\big((Df)(a)\big)(x-a)\Big\|_\infty\leq n\cdot\|(Df)(a)\|_\infty\cdot\|x-a\|_\infty\ .$$

Il s'ensuit immédiatement qu'on a

$$\lim_{x\to a}\Big\|\big((Df)(a)\big)(x-a)\Big\|=n\cdot\|(Df)(a)\|_\infty\cdot 0=0\ .$$

Rédaction "courte" officielle. En utilisant la norme $\|\cdot\|_\infty$ on peut faire le calcul

$$\begin{aligned} 0 &\leq \|f(x)-f(a)\|_\infty \\ &= \left\|\frac{f(x)-f(a)-\big((Df)(a)\big)(x-a)}{\|x-a\|_\infty}\cdot\|x-a\|_\infty+\big((Df)(a)\big)(x-a)\right\|_\infty \\ &\leq \frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|_\infty}{\|x-a\|_\infty}\cdot\|x-a\|_\infty+\Big\|\big((Df)(a)\big)(x-a)\Big\|_\infty \end{aligned}$$

$$\overset{6.1}{\leq} \frac{\big\|f(x)-f(a)-\big((Df)(a)\big)(x-a)\big\|_\infty}{\|x-a\|_\infty}\cdot\|x-a\|_\infty$$
$$+\,n\cdot\|(Df)(a)\|_\infty\cdot\|x-a\|_\infty\ .$$

Le membre de droite tend vers 0 dans la limite $x\to a$ par l'hypothèse que f est différentiable en a, donc par le théorème des gendarmes on en déduit $\lim_{x\to a} f(x)=f(a)$, c'est-à-dire que f est continue en a.

Preuve de 6.5. Comme pour les autres preuves concernant une équivalence entre une propriété vectorielle et la propriété analogue en dimension 1, on procède par double implication, à commencer avec l'implication directe. On suppose donc que f est différentiable en a et on veut montrer que les composantes f_i sont différentiables en a. Avec la définition des composantes par

$$f(x)=\big(f_1(x),\dots,f_p(x)\big)$$

on a donc le schéma "on sait — on veut montrer"

$$\text{on sait}\qquad \lim_{h\to 0}\frac{\big\|f(a+h)-f(a)-\big((Df)(a)\big)(h)\big\|}{\|h\|}=0$$
$$\text{on veut montrer}\qquad \forall i=1,\dots,p\ \exists A_i \text{ une matrice de taille } 1\times n :$$
$$\lim_{h\to 0}\frac{|f_i(a+h)-f_i(a)-A_i(h)|}{\|h\|}=0\ .$$

Mais on nous donne déjà le candidat pour ces matrices A_i : les p lignes de la matrice $(Df)(a)$ de taille $p\times n$. Autrement dit, on désigne par $(Df_i)(a)$ la i-ième ligne de la matrice $(Df)(a)$ et on aura le schéma

$$\text{on sait}\qquad \lim_{h\to 0}\frac{\big\|f(a+h)-f(a)-\big((Df)(a)\big)(h)\big\|}{\|h\|}=0$$
$$\text{on veut montrer}\qquad \forall i=1,\dots,p : \lim_{h\to 0}\frac{|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|}{\|h\|}=0\ .$$

On constate qu'on sait quelque chose avec une norme pour l'instant inconnue, et qu'on veut en déduire quelque chose sur les composantes du même vecteur $f(a+h)-f(a)-\big((Df)(a)\big)(h)$. Choisissons la norme $\|\cdot\|_2$ (bien que, comme dans la preuve de 2.5 les choix $\|\cdot\|_1$ et $\|\cdot\|_\infty$ se déroulent exactement pareille) et remarquons qu'on a pour tout $i=1,\dots,p$ l'inégalité

$$|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|\leq\big\|f(a+h)-f(a)-\big((Df)(a)\big)(h)\big\|_2\ .$$

on a donc pour tout $i=1,\dots,p$ l'encadrement

$$0\leq\frac{|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|}{\|h\|}\leq\frac{\big\|f(a+h)-f(a)-\big((Df)(a)\big)(h)\big\|_2}{\|h\|}\ .$$

Par le théorème des gendarmes et l'hypothèse sur ce qu'on sait, on en déduit qu'on a bien, pour tout $i=1,\dots,p$, l'égalité

$$\lim_{h\to 0}\frac{|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|}{\|h\|}=0\ .$$

- Réciproquement, on a le schéma "on sait — on veut montrer"

$$\text{on sait}\qquad \forall i=1,\dots,p : \lim_{h\to 0}\frac{|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|}{\|h\|}=0$$

on veut montrer $\exists A$ matrice de taille $p \times n$:

$$\lim_{h\to 0} \frac{\big\|f(a+h) - f(a) - A(h)\big\|_2}{\|h\|} = 0 \ .$$

Et de nouveau on nous donne déjà le candidat pour cette matrice A : on empile les p matrices de taille $1 \times n$ en une seule matrice de taille $p \times n$. Regardons de plus près comment cette matrice (empilement de lignes) agit sur un vecteur $h \in \mathbf{R}^n$. On note donc $v = A(h) \in \mathbf{R}^p$ et on veut connaître le lien entre la i-ième composante v_i de v et les différentes lignes dans la matrice. On a donc :

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} = \begin{pmatrix} (Df_1)(a) \\ \vdots \\ (Df_p)(a) \end{pmatrix} \cdot h = \begin{pmatrix} \big((Df_1)(a)\big)(h) \\ \vdots \\ \big((Df_p)(a)\big)(h) \end{pmatrix} \quad \Rightarrow \quad v_i = \big((Df_i)(a)\big)(h) \ .$$

Ceci nous donne la formule (avec la définition de la norme $\|\cdot\|_2$) :

$$\begin{aligned} \frac{\big\|f(a+h) - f(a) - A(h)\big\|_2}{\|h\|} &= \frac{1}{\|h\|} \cdot \sqrt{\sum_{i=1}^p |f_i(a+h) - f_i(a) - \big((Df_i)(a)\big)(h)|^2} \\ &= \sqrt{\sum_{i=1}^p \bigg(\frac{|f_i(a+h) - f_i(a) - \big((Df_i)(a)\big)(h)|}{\|h\|} \bigg)^2} \ . \end{aligned}$$

mais prendre la racine carré, une somme ou un carré sont des opérations continues. On peut donc appliquer le théorème sur la limite d'une composée et conclure qu'on a

$$\begin{aligned} \lim_{h\to 0} & \frac{\big\|f(a+h) - f(a) - A(h)\big\|_2}{\|h\|} \\ &= \lim_{h\to 0} \sqrt{\sum_{i=1}^p \bigg(\frac{|f_i(a+h) - f_i(a) - \big((Df_i)(a)\big)(h)|}{\|h\|} \bigg)^2} \\ &= \sqrt{\sum_{i=1}^p \bigg(\lim_{h\to 0} \frac{|f_i(a+h) - f_i(a) - \big((Df_i)(a)\big)(h)|}{\|h\|} \bigg)^2} \\ &= \sqrt{\sum_{i=1}^p 0^2} = 0 \ . \end{aligned}$$

Si on avait choisi une autre norme $\|\cdot\|_1$ ou $\|\cdot\|_\infty$, le raisonnement serait le même, sauf que dans le cas de la norme $\|\cdot\|_\infty$ il faut utiliser le résultat (qu'on n'a pas montré !) que la limite d'un max (d'un nombre fini de termes) est égale au max des limites.

Rédaction "courte" officielle. On procède par double implication. Si f est différentiable en a, on note $(Df_i)(a)$ la i-ième ligne de la matrice $(Df)(a)$ et on calcule :

$$0 \le \frac{|f_i(a+h) - f_i(a) - \big((Df_i)(a)\big)(h)|}{\|h\|} \le \frac{\big\|f(a+h) - f(a) - \big((Df)(a)\big)(h)\big\|_2}{\|h\|} \ .$$

Par le théorème des gendarmes et l'hypothèse il s'ensuit que f_i est différentiable en a avec "donc" $(Df_i)(a)$ la i-ième ligne de la matrice $(Df)(a)$.

Réciproquement, on note A la matrice de taille $p \times n$ formé en empilant les p matrices $(Df_i)(a)$ de taille $1 \times n$. Avec cela on fait le calcul

$$\begin{aligned}\frac{\big\|f(a+h)-f(a)-A(h)\big\|_2}{\|h\|} &= \frac{1}{\|h\|} \cdot \sqrt{\sum_{i=1}^{p} |f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|^2} \\ &= \sqrt{\sum_{i=1}^{p} \bigg(\frac{|f_i(a+h)-f_i(a)-\big((Df_i)(a)\big)(h)|}{\|h\|}\bigg)^2} \ .\end{aligned}$$

En prenant la limite $h \to 0$ et l'hypothèse, on en déduit que f est différentiable en a avec $(Df)(a) = A$ la matrice formée en empilant les matrices $(Df_i)(a)$.

Preuve de 8.1. Je ferai la preuve dans le cas d'un maximum local ; le cas d'un minimum local se traite de la même façon, la seule différence étant le sens des inégalités. Soit donc $U \subset \mathbf{R}^n$ un ouvert, $a \in U$, $f : U \to \mathbf{R}$ une fonction qui est différentiable en a et supposons qu'il existe un voisinage ouvert $V \subset U$ de a avec la propriété

$$\text{on sait} \qquad \forall x \in V : f(x) \le f(a) \ .$$

Et on demande de montrer qu'on a $(Df)(a) = 0$ en tant que matrice, c'est-à-dire qu'on veut monter

$$\text{on veut} \qquad \forall i = 1, \dots, n : (\partial_i f)(a) = 0 \ .$$

La définition d'une dérivée partielle nous dit qu'on a

$$\begin{aligned}(\partial_i f)(a) &= \lim_{t\to 0} \frac{f(a+te_i)-f(a)}{t} \\ &= \lim_{t\to 0} \frac{f(a_1,\dots,a_{i-1},a_i+t,a_{i+1},\dots,a_n)-f(a_1,\dots,a_n)}{t} \ ,\end{aligned}$$

avec $e_i = (0,\dots,0,1,0,\dots,0)$ la i-ième vecteur de la base canonique de $\mathbf{R}^n$ (avec le 1 au i-ième place). Commençons avec la remarque que l'hypothèse que f est différentiable en a implique (6.9 ou 6.10) que toutes les dérivées partielles $(\partial_i f)(a)$ existent. La seule chose à montrer est donc que sa valeur vaut 0.

Quand on regarde l'expression pour $(\partial_i f)(a)$, on voit la différence $f(a+te_i) - f(a)$ avec $f(a)$ la valeur maximale dans un voisinage de a. Ce numérateur est donc toujours négatif (pas forcément strictement négatif). Mais le dénominateur change de signe selon t négatif ou positif. Pour $t > 0$ le quotient est donc négatif, ce qui implique par le théorème des gendarmes que la limite à droite est négatif. Mais pour $t < 0$ le quotient sera positif, et donc par ce même théorème des gendarmes la limite à gauche sera positif. Mais on sait que la limite existe et en particulier la limite à gauche et la limite à droite donnent la même valeur : la valeur de cette limite. La sule valeur qui est à la fois positif et négatif est 0, ce qui montre qu'on a $(\partial_i f)(a) = 0$ comme souhaité.

Preuve "courte" officielle. Soit donc $U \subset \mathbf{R}^n$ un ouvert, $a \in U$, $f : U \to \mathbf{R}$ une fonction qui est différentiable en a et supposons qu'il existe un voisinage ouvert $V \subset U$ de a avec la propriété

$$\text{on sait} \qquad \forall x \in V : f(x) \le f(a) \ .$$

Soit en plus $i = 1, \ldots, n$ un indice. V étant un voisinage ouvert de a, il existe $r > 0$ tel que $B_r(a) \subset V$, ce qui implique que pour $|t| < r$ l'expression $f(a + te_i)$ est bien définie. Par définition d'un maximum local on a les inégalités

$$-r < t < 0 \Rightarrow \frac{f(a+te_i) - f(a)}{t} \geq 0 \quad \text{et} \quad 0 < t < r \Rightarrow \frac{f(a+te_i) - f(a)}{t} \leq 0 \; .$$

Sachant que $\lim_{t\to 0}\bigl(f(a + te_i) - f(a)\bigr)/t = \ell \in \mathbf{R}$ existe, que si une limite existe, les limites à gauche et à droites existent également avec la même valeur et que une inégalité large est préservée en prenant une limite, on en déduit

$$\ell = \lim_{t\uparrow 0} \frac{f(a+te_i) - f(a)}{t} \geq 0 \qquad \text{et} \qquad \ell = \lim_{t\downarrow 0} \frac{f(a+te_i) - f(a)}{t} \leq 0 \; .$$

On a donc $0 \leq \ell \leq 0$, c'est-à-dire $\ell = 0$ comme voulu.