UNIVERSITE DE RENNES 1

UNIVERSITÉ DE RENNES I
MAÎTRISE DE MATHÉMATIQUES

Travail d'Étude et de Recherche

Différentiabilité des fonctions lipschitziennes

effectué par Florian Caro et Alexandre Popier
sous la responsabilité de Monsieur J. Camus

mai 1999

Table des matières

Introduction

Au siècle dernier s'est posée la question de savoir si les fonctions appartenant à telle ou telle catégorie, comme les fonctions continues ou monotones, admettent nécessairement des dérivées (sur $\mathbb{R}$ ou sur $\mathbb{R}$ privé d'un ensemble de mesure nulle). C'est Weierstrass qui a fourni le premier résultat en construisant une *fonction continue sans dérivée*. En voici un exemple dû à Van der Waerden, fondé sur le fait qu'une suite de nombres entiers ne peut être convergente que si ses termes restent constants à partir d'un certain rang. Pour $x \in \mathbb{R}$, on pose :

$$f(x) = \sum_{n=0}^{\infty} \frac{\{10^n x\}}{10^n}$$

où $\{x\}$ désigne la distance de x à l'entier le plus proche. f est continue sur $\mathbb{R}$, mais dérivable nulle part (cf. Annexe 1 pour la démonstration de ce résultat).

Donc l'hypothèse de continuité n'est pas suffisante. Par contre on va montrer dans ce travail qu'une fonction lipschitzienne est dérivable sur $\mathbb{R}$, sauf éventuellement sur un ensemble de mesure nulle. Pour cela on commencera par établir un résultat analogue sur les fonctions monotones définies sur $\mathbb{R}$, d'où on déduira le théorème sur les fonctions lipschitziennes. Dans une seconde partie, on l'étendra aux fonctions définies sur $\mathbb{R}^n$.

On notera $\mathcal{L}^n(\mathrm{A})$ la mesure de Lebesgue de l'ensemble A inclus dans $\mathbb{R}^n$.

1 Un théorème de Lebesgue

1.1 Fonctions monotones et dérivabilité

Dans ce paragraphe on veut montrer le théorème suivant :

Théorème 1

Toute fonction monotone $f : [a,b] \to \mathbb{R}$ est dérivable presque partout.

Pour cela on va l'établir d'abord dans le cas où f est monotone et continue, puis on va voir ce qui change si f est supposée seulement monotone.

Lemme 1

Soit g une fonction continue définie sur $[a,b]$ et soit

$$E = \{x \in \,]a,b[\; / \;\exists\; \xi > x \;/\; g(x) < g(\xi)\}$$

Alors soit E est vide, soit

$$E = \bigcup_{k=0}^{\infty}]a_k, b_k[$$

avec union disjointe, et avec $g(a_k) \leq g(b_k)$ pour tout k.

Démonstration :
E est ouvert car si x_0 appartient à E il existe $\xi_0 > x_0$ tel que $g(x_0) < g(\xi_0)$. Donc si x est voisin de x_0 on aura $g(x) < g(\xi_0)$ par continuité de g. Donc E peut s'écrire comme la réunion disjointe d'intervalles $]a_k, b_k[$.

Reste à montrer que $g(a_k) \leq g(b_k)$ pour tout k.
Soit x appartenant à $]a_k, b_k[$ et $x_1 =$ sup { x $\leq$ t $\leq b_k$ / $g(t) \geq g(x)\}$. On veut montrer que $x_1 = b_k$. Supposons $x_1 \neq b_k$. Comme x_1 appartient à E, il existe $\xi_1 > x_1$ tel que $g(x_1) < g(\xi_1)$. Par définition du sup, on a $\xi_1 > b_k$ et comme b_k n'appartient pas à E, on a : $g(\xi_1) \leq g(b_k)$. De plus $g(b_k) < g(x_1)$ sinon $g(b_k) \geq g(x_1) \geq g(x)$, d'où, par définition de x_1, $x_1 = b_k$, ce qui est contraire à notre hypothèse. Finalement :

$$g(x_1) < g(\xi_1) \leq g(b_k) < g(x_1)$$

Donc $x_1 = b_k$ et $g(x) \leq g(b_k)$. Pour x tendant vers a_k, on obtient $g(a_k) \leq g(b_k)$.
□

Théorème 2
Si f : [a,b] → $\mathbb{R}$ *est une fonction continue monotone, alors f est dérivable presque partout.*

Tout d'abord on va introduire $\Delta(x,h) = \frac{f(x+h)-f(x)}{h}$, pour tout $x \in [a,b]$ et tout $h \neq 0$, et les nombres suivants :

$$\Lambda_d(x) = \lim_{h \to 0, h>0} sup\{\Delta(x,\epsilon), 0 < \epsilon < h\}$$

$$\lambda_d(x) = \lim_{h \to 0, h>0} inf\{\Delta(x,\epsilon), 0 < \epsilon < h\}$$

$$\Lambda_g(x) = \lim_{h \to 0, h<0} sup\{\Delta(x,\epsilon), h < \epsilon < 0\}$$

$$\lambda_g(x) = \lim_{h \to 0, h<0} inf\{\Delta(x,\epsilon), h < \epsilon < 0\}$$

Définition 1

Les nombres précédents sont appelés *nombres dérivés de f en x.*

Pour montrer que f est dérivable presque partout, il suffit de démontrer que, pour presque tout x :

1. $\Lambda_d(x) < \infty$
2. $\Lambda_d(x) \leq \lambda_g(x)$

En effet en appliquant (2) à $x \mapsto -f(-x)$, on aura : $\Lambda_g(x) \leq \lambda_d(x)$ presque partout. Donc : $0 \leq \Lambda_d(x) \leq \lambda_g(x) \leq \Lambda_g(x) \leq \lambda_d(x) \leq \Lambda_d(x) < \infty$ presque partout.

Démonstration :
On suppose f croissante.

Montrons 1. $\Lambda_d(x) < \infty$ pour presque tout x.
Soit $E_\infty = \{x \in [a,b], \Lambda_d(x) = \infty\}$. Pour tout $c \leq 0$, on pose $E_c = \{x, c < \Lambda_d(x)\}$. Par définition pour tout c, $E_\infty \subseteq E_c$. Or si $c < \Lambda_d(x)$, il existe $\xi > x$ tel que

$$\frac{f(\xi) - f(x)}{\xi - x} > c$$

d'où si $g(x) = f(x) - cx$, $g(\xi) > g(x)$. Ainsi en appliquant le lemme on obtient :

$$E_c = \bigcup_{k=0}^{\infty}]a_k, b_k[$$

avec $f(b_k) - cb_k \leq f(a_k) - ca_k$, d'où $c(b_k - a_k) \leq f(b_k) - f(a_k)$. Donc

$$c\sum_{k=0}^{\infty}(b_k - a_k) \leq \sum_{k=0}^{\infty}(f(b_k) - f(a_k)) \leq f(b) - f(a)$$

Donc pour tout k entier

$$b_k - a_k \leq \frac{f(b) - f(a)}{c}$$

Ceci montre que, pour c suffisamment grand, la longueur totale des intervalles (a_k,b_k) sera aussi petite qu'on veut. Donc $\mathcal{L}^1$ $(E_\infty) = 0$.

Montrons 2. $\Lambda_d(x) \leq \lambda_g(x)$ pour presque tout x.
Soient $0 < c < C$ arbitraires. Soit $g_1(x) = f(-x) + cx$ et

$$E_1 = -\{x \in [a,b] / \exists \xi > x / g_1(\xi) > g_1(x)\}$$

g_1 est continue, donc d'après le lemme, on peut décomposer

$$-E_1 = \bigcup_{k=0}^{\infty}] - b_k, -a_k[$$

avec $a_k < b_k$ et $g_1(-b_k) \leq g_1(-a_k)$. Soit

$$m_1 = \mathcal{L}^1(E_1) = \sum_{k=0}^{\infty}(b_k - a_k)$$

On a pour tout k : $f(b_k) - cb_k \leq f(a_k) - ca_k$. Donc

$$\sum_{k=0}^{\infty}(f(b_k) - f(a_k)) \leq cm_1$$

Soit $g_2(x) = f(x) - Cx$ et soit $E_{2,k} = \{x \in [a_k, b_k] \ / \ \exists \ \xi > x/g_2(\xi) > g_2(x)\}$. g_2 est continue, donc d'après le lemme :

$$E_{2,k} = \bigcup_{l=0}^{\infty}]a_{k,l}, b_{k,l}[$$

avec $a_{k,l} < b_{k,l}$ et $g(a_{k,l}) \leq g(b_{k,l})$.

Soit

$$m_2 = \mathcal{L}^1(\bigcup_{k=0}^{\infty} E_{2,k}) = \sum_{k,l}(b_{k,l} - a_{k,l})$$

On a pour tout k et pour tout l, $f(a_{k,l}) - Ca_{k,l} \leq f(b_{k,l}) - Cb_{k,l}$. Donc

$$Cm_2 \leq \sum_{k,l}(f(b_{k,l}) - f(a_{k,l})) \leq \sum_{k=0}^{\infty}(f(b_k) - f(a_k)) \leq cm_1$$

par croissance de f. Donc $m_2 \leq \alpha m_1$ avec $\alpha = \frac{c}{C} < 1$.
En réitérant ce procédé, on obtient une suite m_n telle que $m_{2n} \leq \alpha \ m_{2n-1} \leq \alpha^n m_1$ qui tend vers 0 quand n tend vers $+\infty$.
Or pour tout n entier,

$$E_{cC} = \{x/\Lambda_d(x) > C, \lambda_g(x) < c\} \subseteq E_n$$

En effet, si $x \in E_{cC}$, alors il existe $\xi < x$ tel que

$$\frac{f(\xi) - f(x)}{\xi - x} < c$$

Donc -$\xi > -x$ et $f(-(-\xi)) - f(-(-x)) > c(\xi - x)$, ie $g_1(-\xi) > g_1(-x)$, donc $x \in E_1$.
De même il existe $\xi' > x$ tel que

$$\frac{f(\xi') - f(x)}{\xi' - x} > C$$

Donc $g_2(\xi') =$ f(ξ') - C$\xi' > g_2(x) = f(x) - Cx$, donc $x \in E_2$.
Si x est tel que $\Lambda_d(x) > \lambda_g(x)$, on peut choisir c et C rationnels tels que : $x \in E_{cC}$. Donc

$$\{x, \Lambda_d(x) > \lambda_g(x)\} \subseteq \bigcup_{denombrable} E_{cC}$$

Ainsi $\mathcal{L}^1(\{x, \Lambda_d(x) > \lambda_g(x)\}) = 0$.

□

On a donc montré qu'une fonction monotone continue est dérivable presque partout.

Théorème 3

Soient D une partie de $\mathbb{R}$, $f : D \rightarrow \mathbb{R}$ une application monotone, a un point de $\overline{\mathbb{R}}$ tel que a soit adhérent à $D \cap]a, +\infty[$ (resp. $D \cap] - \infty, a[$). Alors f admet une limite (finie ou infinie) à droite (resp. à gauche) au point a.

Démonstration :
Supposons f croissante et a $\in \overline{D \cap]a, +\infty[}$. On pose

$$l = \inf_{t \in D \ \cap \ ]a,+\infty[} f(t)$$

On a : $l \in \overline{\mathbb{R}}$ et :

$$\forall l' \in]l, +\infty] \ \exists \ t_0 \in D \cap]a, +\infty[\ / \ l \leq f(t_0) < l'$$

Donc

$$\forall t \in D \cap \ ]a, t_0[\quad l \leq f(t) \leq f(t_0) < l'$$

d'où

$$l = \lim_{\substack{t \rightarrow a \\ t > a}} f(t) = f(a+0)$$

□

Corollaire 1

Soit $f : D \rightarrow \mathbb{R}$ croissante et a un point adhérent à $D \cap]a, +\infty[$. Alors f admet une limite finie à droite au point a si et seulement si f est minorée sur $D \cap]a, +\infty[$.

Démonstration :
On a :

$$\lim_{\substack{t \rightarrow a \\ t > a}} f(t) = l = \inf_{t \in D \ \cap \ ]a,+\infty[} f(t)$$

et $l \in \overline{\mathbb{R}}$.
Or $D \cap \ ]a, +\infty[$ est non vide (car a est adhérent à cet ensemble), donc $l < +\infty$.
Donc l est finie si et seulement si

$$-\infty < \inf_{t \in D \ \cap \ ]a,+\infty[} f(t)$$

si et seulement si f est minorée sur $D \cap]a, +\infty[$.

□

De là on a immédiatement le corollaire suivant :

Corollaire 2

Soient I un intervalle de $\mathbb{R}$, $f : I \to \mathbb{R}$ une application monotone et a un point de I. Si $a \neq sup(I)$, f admet une limite à droite (finie) f(a+0) et si $a \neq inf(I)$, f admet une limite à gauche (finie) f(a-0).

On en déduit le théorème suivant :

Théorème 4

Soit f une application monotone d'un intervalle réel I dans $\mathbb{R}$. L'ensemble E des points de discontinuité de f est au plus dénombrable.

Démonstration :
On suppose f croissante et I de la forme (a,b). On définit la fonction σ sur I par :
$\sigma(t) = f(t+0) - f(t-0)$, pour $t \in]a, b[$
et, s'il y a lieu, $\sigma(a) = f(a+0) - f(a)$ et $\sigma(b) = f(b) - f(b-0)$.
On a les propriétés suivantes : $\forall t \in]a, b[\ \sigma(t) \geq 0$, et $E = \{t \in I \ / \ \sigma(t) > 0\}$.
Construisons $r : E \to \mathbb{R}$, avec pour $r(t)$, $t \in E$, un rationnel tel que $r(t) \in]f(t-0), f(t+0)[$ si $t \in]a, b[$ et éventuellement $r(a) \in]f(a), f(a+0)[$ et $r(b) \in]f(b-0), f(b)[$.
Par construction, r est une application injective de E dans $\mathbb{Q}$. Donc E est dénombrable.
$\square$

De là la preuve du théorème 1 se déduit aisément.

Démonstration du théorème 1 :
Soit, d'après le lemme, $E = \{x_i / i \in \mathbb{N}\}$ l'ensemble des points de discontinuité de f, avec $x_0 < x_1 < ... < x_n < ....$ Pour tout i, f est continue sur $]x_{i-1}, x_i[$. Soit f_i définie sur $[x_{i-1}, x_i]$ par f_i(x) = f(x) si x $\in]x_{i-1}, x_i[$ et $f_i(x_{i-1})$ = f(x_{i-1} + 0) et $f_i(x_i)$ = f(x_i - 0). Alors f_i est continue monotone sur $[x_{i-1}, x_i]$. Donc il existe $E_i \subseteq [x_{i-1}, x_i]$ négligeable tel que sur $[x_{i-1}, x_i] \setminus E_i$, f_i est dérivable. Donc f est dérivable sur

$$[a, b] \setminus \{(\bigcup_i E_i) \cup E\}$$

et

$$(\bigcup_i E_i) \cup E$$

est de mesure nulle.
$\square$

1.2 Fonctions à variations bornées

Ici on va étudier une classe plus importante que celle des fonctions lipschitziennes : les fonctions à variations bornées, et établir un théorème analogue à celui sur les fonctions monotones.

Définition 2

On appelle <u>fonction à variations bornées</u> une fonction f : [a,b] → $\mathbb{R}$ telle qu'il existe $M > 0$ tel que pour toute subdivision $\sigma = (x_i)_{i=0,n}$ de [a,b]

$$\Sigma_{ab} = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq M$$

On pose T(a,b) = Inf M, la variation totale de f sur l'intervalle [a,b]. La variation totale est une fonction "additive de l'intervalle", ce qui veut dire que si c est un point de l'intervalle [a,b], la fonction f est à variations bornées sur [a,b] si et seulement si elle l'est sur [a,c] et sur [c,b], et qu'alors T(a,b) = T(a,c) + T(c,b). Il suffit d'observer que Σ_{ab} ne peut qu'augmenter si on ajoute un nouveau point de décomposition.

De plus si f = $f_1 - f_2$ où f_1 et f_2 sont deux fonctions croissantes, on a

$$\Sigma_{ab} \leq f_1(b) - f_1(a) + f_2(b) - f_2(a)$$

Donc f est à variations bornées. La réciproque est vraie et est due à Camille Jordan.

Théorème 5

Toute fonction à variations bornées est la différence de deux fonctions croissantes

Démonstration :

On introduit pour x appartenant à [a,b] T(x) = T(a,x) et on va montrer que les fonctions T : x ↦ T(x) et x ↦ T(x) - f(x) sont croissantes.
En effet si $\xi > $ x, on a : T(a,ξ) = T(a,x) + T(x,ξ), d'où T(ξ) - T(x) = T(x,ξ) ≥ 0.
Pour montrer que T(x) - f(x) $\leq$ T(ξ) - f(ξ), il suffit de montrer que f(ξ) - f(x) $\leq$ T(x,ξ). Or |f(ξ) - f(x)| est une décomposition particulière, donc par définition, f(ξ) - f(x) $\leq$ T(x,ξ).
□

Une autre décomposition possible de f est

$$f(x) = \frac{1}{2}(T(x) + f(x)) - \frac{1}{2}(T(x) - f(x))$$

A partir de ce théorème et du théorème 1, on a le théorème de Lebesgue suivant.

Théorème 6

Toute fonction à variations bornées est dérivable presque partout

Définition 3

Soit I un intervalle de $\mathbb{R}$. Une fonction $f : I \rightarrow \mathbb{R}$ est dite <u>lipschitzienne</u> si pour une certaine constante C, pour tout (x,y) $\in I^2$, on a : |f(x) - f(y)| $\leq$ C|x - y|. La plus petite constante C vérifiant l'inégalité précédente, est notée

$$Lip(f) \equiv sup\{\frac{|f(x) - f(y)|}{|x - y|}, x, y \in I, x \neq y\}$$

Définition 4

Soit I un intervalle de $\mathbb{R}$. Une fonction $f : I \to \mathbb{R}$ est dite localement lipschitzienne si pour tout compact $K \subseteq I$, il existe une constante C_K telle que

$$|f(x) - f(y)| \leq C_K |x - y|$$

pour tout $x,y \in K$.

Enfin il est facile de voir qu'une fonction lipschitzienne est à variations bornées. Donc on va finir ce paragraphe par le théorème suivant :

Théorème 7

Toute fonction lipschitzienne de $[a,b]$ dans $\mathbb{R}$ est dérivable presque partout

2 Théorème de Rademacher

Dans ce paragraphe, on va utiliser le théorème 7, pour étendre ce résultat aux fonctions définies sur $\mathbb{R}^n$.

2.1 Définitions et extensions des fonctions lipschitziennes

On commence par rappeller les définitions de *lipschitzienne* et de *différentiable*. Puis on va démontrer comment prolonger les fonctions lipschitziennes d'un sous-ensemble de $\mathbb{R}^n$ à $\mathbb{R}^n$ tout entier.

Définition 5

Soit $A \subseteq \mathbb{R}^n$. Une fonction $f : A \to \mathbb{R}^m$ est dite lipschitzienne si pour une certaine constante C, pour tout (x,y) $\in A^2$, on a : $|f(x) - f(y)| \leq C|x - y|$. La plus petite constante C vérifiant l'inégalité précédente, est notée

$$Lip(f) \equiv sup\{\frac{|f(x)-f(y)|}{|x-y|}, x, y \in A, x \neq y\}$$

Définition 6

Soit $A \subseteq \mathbb{R}^n$. Une fonction $f : A \to \mathbb{R}^m$ est dite localement lipschitzienne si pour tout compact $K \subseteq A$, il existe une constante C_K telle que

$$|f(x) - f(y)| \leq C_K|x - y|$$

pour tout x,y $\in K$.

Définition 7

Une fonction $f : \mathbb{R}^n \to \mathbb{R}^m$ est différentiable en $x \in \mathbb{R}^n$ s'il existe une application linéaire $L : \mathbb{R}^n \to \mathbb{R}^m$ telle que

$$\lim_{y \to x} \frac{|f(y) - f(x) - L(x-y)|}{|x-y|} = 0$$

ou de manière équivalente : $f(y) = f(x) + L(y\text{-}x) + o(|y\text{-}x|)$ quand $y \to x$.

Si une telle application linéaire existe, elle est unique et on la note Df(x), qui est la différentielle de f en x.

Théorème 8

Soit $A \subseteq \mathbb{R}^n$, et soit $f : A \to \mathbb{R}^m$ lipschitzienne. Alors il existe une fonction lipschitzienne $\overline{f} : \mathbb{R}^n \to \mathbb{R}^m$ telle que :
i. $\overline{f} = f$ sur A
ii. $Lip(\overline{f}) \leq \sqrt{m}\ Lip(f)$

Démonstration :
1. Supposons d'abord $f : A \to \mathbb{R}$. On définit :

$$\overline{f}(x) \equiv \inf_{a \in A}\{f(a) + Lip(f)|x-a|\}$$

Si b $\in$ A, on a $\overline{f}$(b) = f(b). En effet il est immédiat que $\overline{f}$(b) $\leq$ f(b). De plus, pour tout a $\in$ A, f(a) + Lip(f)|b-a| $\geq$ f(b), d'où $\overline{f}$(b) $\geq$ f(b).
Si x,y sont dans $\mathbb{R}^n$, alors

$$\overline{f}(x) \leq \inf_{a \in A}\{f(a) + Lip(f)(|y-a| + |x-y|)\} = \overline{f}(y) + Lip(f)|x-y|$$

et de même

$$\overline{f}(y) \leq \overline{f}(x) + Lip(f)|x-y|$$

2. Dans le cas général, $f : A \to \mathbb{R}^m$, $f = (f_1,...,f_m)$, on définit $\overline{f} = (\overline{f_1},...,\overline{f_m})$. Alors

$$|\overline{f}(x) - \overline{f}(y)|^2 = \sum_{i=1}^{m} |\overline{f_i}(x) - \overline{f_i}(y)|^2 \leq m(Lip(f))^2|x-y|^2$$

□

A partir de ce théorème, on peut supposer qu'une fonction lipschitzienne est toujours définie sur $\mathbb{R}^n$ tout entier.

2.2 Le théorème de Rademacher

Ce théorème est une extension du théorème 7 :

Théorème 9

Soit f : $\mathbb{R}^n \to \mathbb{R}^m$ une fonction localement lipschitzienne. Alors f est différentiable presque partout.

Preuve :
1. On peut supposer que m = 1. De plus la différentiabilité étant une notion locale, on peut supposer que f est lipschitzienne sur un voisinage d'un point de $\mathbb{R}^n$ puis l'étendre sur $\mathbb{R}^n$, d'après le théorème 8.
2. Fixons un vecteur v appartenant à $\mathbb{R}^n$ avec |x| = 1 et on définit, pour x $\in$ $\mathbb{R}^n$:

$$D_v f(x) \equiv \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$$

pourvu que cette limite existe.
3. Montrons que $D_v f(x)$ existe pour presque tout x dans $\mathbb{R}^n$ (assertion 1).
En effet, f étant continue,

$$\overline{D_v f(x)} \equiv \limsup_{t \to 0} \frac{f(x+tv) - f(x)}{t} = \lim_{k \to \infty} \sup_{\substack{0 < |t| < 1/k \\ t \ rationnel}} \frac{f(x+tv) - f(x)}{t}$$

est mesurable, ainsi que

$$\underline{D_v f(x)} = \liminf_{t \to 0} \frac{f(x+tv)-f(x)}{t}$$

Donc

$$A_v \equiv \{x \in \mathbb{R}^n / D_v f(x) \ n' \ existe \ pas\} = \{x \in \mathbb{R}^n / \underline{D_v f(x)} < \overline{D_v f(x)}\}$$

est Lebesgue mesurable.
Maintenant pour tout x,v dans $\mathbb{R}^n$ avec |v| $= 1$, on définit $\phi : \mathbb{R} \to \mathbb{R}$ par $\phi(t) \ = \ f(x+tv)$. Alors ϕ est lipschitzienne, donc dérivable presque partout sur $\mathbb{R}$, d'après le théorème de Lebesgue (théorème 7).
Donc $\mathfrak{L}^1$ ($A_v \cap$ L) vaut 0 pour toute ligne L parallèle à v. Le théorème de Fubini permet de conclure que $\mathfrak{L}^1(A_v) = 0$. En effet on commence par compléter (v) en une base de $\mathbb{R}^n$. Alors

$$\mathfrak{L}^n(A_v) \ = \ \int_{\mathbb{R}^n} \chi_{A_v}(x)dx = \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}.v} \chi_{A_v}(x_1,...,x_n)dx_n \right) dx_1...dx_{n-1}$$

en utilisant le théorème de Fubini.
Donc

$$\mathfrak{L}^n(A_v) \ = \ \int_{\mathbb{R}^{n-1}} \mathfrak{L}^1(A_v \cap L(x_1,...,x_{n-1}))dx_1...dx_{n-1}$$

où $L(x_1,...,x_{n-1})$ est une droite parallèle à v et passant par $(x_1,...,x_{n-1})$.

4. L'assertion (1) implique que

$$grad \ f(x) = \left(\frac{\partial f}{\partial x_1}(x), ..., \frac{\partial f}{\partial x_n}(x) \right)$$

existe pour presque tout x.

5. Montrons maintenant que $D_v f(x) = v.grad \ f(x)$ pour presque tout x (assertion 2). Soit $\zeta \in C_0^\infty(\mathbb{R}^n)$ (ie C^∞ à support compact).
Alors

$$\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t} \ \zeta(x)dx = -\int_{\mathbb{R}^n} f(x)\frac{\zeta(x)-\zeta(x-tv)}{t}dx$$

Or, si t $= \frac{1}{k}$, on a

$$|\frac{f(x+\frac{v}{k})-f(x)}{\frac{1}{k}} \ \zeta(x)| \leq Lip(f)|v||\zeta(x)|$$

et |v| $= 1$. De plus ζ est dans $\mathbb{L}^1$. Donc, d'après le théorème de convergence dominée, on a :

$$\int_{\mathbb{R}^n} D_v f(x)\zeta(x)dx = -\int_{\mathbb{R}^n} f(x)D_v\zeta(x)dx = -\sum_{i=1}^n v_i \int_{\mathbb{R}^n} f(x)\frac{\partial \zeta}{\partial x_i}(x)dx$$

$$= \sum_{i=1}^n v_i \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(x)\zeta(x)dx = \int_{\mathbb{R}^n} (v.grad \ f(x))\zeta(x)dx$$

en utilisant le théorème de Fubini et par intégration par parties, justifiée dans le paragraphe 2.3.

Comme l'égalité est vérifiée pour tout $\zeta \in C_0^\infty(\mathbb{R}^n)$, on a $D_v f(x) = v.grad\ f(x)$ presque partout (justification de ce résultat : cf Annexe 2).

6. Soit $(v_k)_{k\in\mathbb{N}}$ une suite dense dans ∂ B(0,1). Soient :

$$A_k = \{x \in \mathbb{R}^n / D_{v_k} f(x), grad\ f(x)\ existent\ et\ D_{v_k} f(x) = v_k.grad\ f(x)\}$$

$$A = \bigcap_{k\in\mathbb{N}} A_k$$

Alors $\mathcal{L}^n(\mathbb{R}^n \setminus A) = 0$. En effet,

$$\mathcal{L}^n(\mathbb{R}^n \setminus A) = \mathcal{L}^n(A^c) = \mathcal{L}^n(\bigcup_{k\in\mathbb{N}} A_k^c) \leq \sum_{k=0}^{\infty} \mathcal{L}^n(A_k^c) = 0$$

Montrons que pour tout x appartenant à A, f est différentiable en x (assertion 3).
Soit x $\in$ A fixé, v $\in \partial$ B(0,1) et t $\in \mathbb{R}$ non nul. Notons :

$$Q(x,v,t) = \frac{f(x+tv) - f(x)}{t} - v.grad\ f(x)$$

Alors pour tout v' $\in \partial$ B(0,1),

$$\begin{aligned} |Q(x,v,t) - Q(x,v',t)| &\leq Lip(f)|v-v'| + |grad\ f(x)||v-v'| \\ &\leq (1+\sqrt{n})Lip(f)|v-v'| \quad (*) \end{aligned}$$

Soit $\epsilon > 0$ fixé. On a :

$$\partial B(0,1) \subseteq \bigcup_{k\in\mathbb{N}} B(v_k, \epsilon)$$

Par compacité de ∂ B(0,1), il existe N $\in \mathbb{N}$ tel que

$$\partial B(0,1) \subseteq \bigcup_{k=0}^{N} B(v_k, \epsilon)$$

Donc si v $\in \partial$ B(0,1), $|v - v_k| \leq \epsilon$ pour au moins un k $\in \{0, ..., N\}$. Mais aussi x $\in A_k$ pour tout k, donc Q(x,v_k,t) a une limite quand t tend vers 0 et la limite est nulle, pour tout k.

Donc il existe $\delta = \delta(x,\epsilon) > 0$ tel que si $0 <$ |t| $< \delta$, alors |Q(x,v_k,t)| $\leq \epsilon$ (**) pour tout k $\in \{0, ..., N\}$.
Ainsi pour tout v $\in \partial$ B(0,1), il existe k $\in \{0, ..., N\}$ tel que

$$|Q(x,v,t)| \leq |Q(x,v_k,t)| + |Q(x,v,t) - Q(x,v_k,t)| \leq \epsilon(1 + (1+\sqrt{n})Lip(f))$$

si $0 <$ |t| $< \delta$, avec les inégalités (*) et (**).

Ceci signifie que Q(x,v,t) tend vers 0 quand t tend vers 0, indépendamment de v.
Soit y $\in \mathbb{R}^n$, y $\neq$ x. On écrit y = x + tv, avec |v| = 1, t = |y-x|. Alors

$$f(y) - f(x) - gradf(x).(y-x) = f(x+tv) - f(x) - tv.grad\ f(x) = o(t) = o(|y-x|)$$

pour y tendant vers x (car δ ne dépend pas de v).
Donc f est différentiable sur A.

$\square$

2.3 Intégrale de Lebesgue et intégration par parties

L'objet de ce paragraphe est de justifier l'intégration par parties que l'on a effectuée dans la démonstration du théorème de Rademacher. En effet celle-ci ne respecte pas a priori les hypothèses habituelles de l'intégrale de Riemann. Il existe un théorème plus général pour l'intégrale de Lebesgue (voir [4], p.199, proposition 5).

Dans notre cas, on a une fonction ζ dans $C_0^\infty(\mathbb{R}^n)$ et une fonction f lipschitzienne dont la dérivée par rapport à x_i $\frac{\partial f}{\partial x_i}$ existe presque partout et est dans $\mathcal{L}^\infty$. On a alors :

$$\int_{\mathbb{R}} \frac{\partial f}{\partial x_i}(x)\zeta(x)dx_i = -\int_{\mathbb{R}} f(x)\frac{\partial \zeta}{\partial x_i}(x)dx_i$$

Lemme 2 *Soit ζ dans $C_0^\infty(\mathbb{R})$ et f lipschitzienne dont la dérivée existe presque partout et est dans $\mathcal{L}^\infty$. Alors on a :*

$$\int_{\mathbb{R}} f'(x)\zeta(x)dx = -\int_{\mathbb{R}} f(x)\zeta'(x)dx$$

Démonstration :
En effet soit $(j_\epsilon)_{\epsilon>0}$ une famille régularisante, ie une famille de fonctions telles que :

$$\begin{aligned} &i. j_\epsilon \in C_0^\infty(\mathbb{R}) \\ &ii. j_\epsilon \geq 0 \\ &iii. \textstyle\int_{\mathbb{R}} j_\epsilon(x)dx \ = \ 1 \\ &iv.\ supp(j_\epsilon) \ = \ B_f(0,\epsilon) \end{aligned}$$

Soit Φ une fonction dans $C_0^\infty(\mathbb{R})$ valant 1 dans un voisinage du support de ζ. On pose $f_\epsilon = (\Phi f) \star j_\epsilon$ qui est ainsi dans $C^\infty(\mathbb{R})$. Comme $\Phi f \in \mathcal{L}^1$, f_ϵ tend vers Φf dans $\mathcal{L}^1$.

De plus $f'_\epsilon = (\Phi f)' \star j_\epsilon$ car :

$$\frac{f_\epsilon(x+h) - f_\epsilon(x)}{h} = \int_{\mathbb{R}} j_\epsilon(y)\frac{(\Phi f)(x-y+h) - (\Phi f)(x-y)}{h}dy$$

et on a :

$$|j_\epsilon(y)\frac{(\Phi f)(x-y+h) - (\Phi f)(x-y)}{h}|$$
$$\leq j_\epsilon(y)(|\frac{\Phi(x-y+h)(f(x-y+h) - f(x-y))}{h}| + |\frac{f(x-y)(\Phi(x-y+h) - \Phi(x-y))}{h}|)$$
$$\leq j_\epsilon(y)(Lip(f)sup(|\Phi|) + sup(|\Phi'|)|f(x-y)|) \in \mathcal{L}^1$$

Par le théorème de convergence dominée en faisant tendre h vers 0, on obtient le résultat annoncé.

Enfin comme $(\Phi f)'$ est dans $\mathcal{L}^1$, f'_ϵ tend vers $(\Phi f)'$ dans $\mathcal{L}^1$.

On a (intégration par parties avec intégrale de Riemann) :

$$\int_{\mathbb{R}} f'_\epsilon(x)\zeta(x)dx = -\int_{\mathbb{R}} f_\epsilon(x)\zeta'(x)dx$$

Le membre de droite tend vers :

$$\int_{\mathbb{R}} (\Phi f)(x)\zeta'(x)dx \;=\; \int_{\mathbb{R}} f(x)\zeta'(x)dx$$

Le membre de gauche tend vers :

$$\int_{\mathbb{R}} (\Phi f)'(x)\zeta(x)dx \;=\; \int_{\mathbb{R}} (\Phi' f)(x)\zeta(x)dx + \int_{\mathbb{R}} (\Phi f')(x)\zeta(x)dx \;=\; 0 + \int_{\mathbb{R}} f'(x)\zeta(x)dx$$

d'où l'égalité voulue.

$\square$

Conclusion

On a ainsi montré qu'une fonction continue peut être dérivable en aucun point. Par contre en ajoutant l'hypothèse "fonction lipschitzienne" (donc en réduisant la classe de fonctions), la fonction devient dérivable presque partout (différentiable dans le cas de $\mathbb{R}^n$). Dans le même esprit, on pourrait s'intéresser à une classe encore plus petite de fonctions continues : les fonctions convexes. En effet celles-ci sont localement lipschitziennes, donc différentiables presque partout. Mais en plus on peut montrer qu'elles sont deux fois différentiables presque partout (cf. [2]).

Annexe 1

Weierstrass avait introduit la fonction suivante :

$$W(t) \ = \ \sum_{n=0}^{\infty} \omega^{-nH} cos(\omega^n t)$$

où $\omega > 1$ et $0 < H < 1$. Ici on va plutôt s'intéresser à la fonction :

$$f(x) = \sum_{n=0}^{\infty} \frac{\{10^n x\}}{10^n}$$

qui elle aussi est continue sur $\mathbb{R}$, mais dérivable en aucun point.

i. En effet, la fonction $x \mapsto \{x\}$ est continue sur $\mathbb{R}$ (1-périodique). Et, de plus pour tout $n \in \mathbb{N}$ et tout $x \in \mathbb{R}$, ona :

$$|\frac{\{10^n x\}}{10^n}| \leq \frac{1}{10^n}$$

Donc f est continue sur $\mathbb{R}$.

ii. Soit $0 \leq x < 1$. Alors on peut écrire x sous la forme $x \ = \ 0, a_1 a_2 \ldots a_n \ldots$, quitte à ajouter des zéros si nécessaire. Deux cas sont à distinguer : si $0, a_{n+1} a_{n+2} \ldots \leq \frac{1}{2}$, alors $\{10^n x\} \ = \ 0, a_{n+1} a_{n+2} \ldots$, sinon $\{10^n x\} \ = \ 1 - 0, a_{n+1} a_{n+2} \ldots$.

On définit alors la suite (h_m) de la manière suivante : si $a_m = 4$ ou 9 on pose $h_m = -10^{-m}$ et $h_m = 10^{-m}$ sinon. On a alors :

$$\frac{f(x + h_m) - f(x)}{h_m} \ = \ \sum_{n=0}^{+\infty} \pm \frac{\{10^n (x \pm 10^{-m})\} - \{10^n x\}}{10^{n-m}} \ = \ \sum_{n=0}^{+\infty} b_{n,m}$$

Pour $n \geq m$, $b_{n,m} = 0$ et pour $n < m$, $b_{n,m} = \pm 1$. Donc la série précédente est égale à un entier pair ou impair suivant la parité de m. Donc ce n'est pas une série convergente. Comme la suite (h_m) est une suite convergente vers 0, on en déduit que f n'est pas dérivable en x pour tout $x \in [0, 1[$, donc pour tout $x \in \mathbb{R}$.

A la fin de ce rapport, on a tracé (sous Maple) les sommes partielles (jusqu'à l'entier n) des fonctions suivantes :

$$f_a(x) \ = \ \sum_{k=0}^{\infty} \frac{a^k \{10^k x\}}{10^k}$$

où a est un réel appartenant à l'intervalle [1,10[. D'après la démonstration précédente, f_a est continue sur $\mathbb{R}$ et dérivable en aucun point. De plus le graphe de f_a est une courbe dont la dimension fractale est donnée par la formule $d \ = \ 1 + \frac{ln(a)}{ln(10)}$.

Sur les fonctions continues, non dérivables et de leur dimension fractale, voir [5].

Annexe 2

Le but ici est de justifier le passage suivant :

$$\int_{\mathbb{R}^n} D_v f(x)\zeta(x)dx = \int_{\mathbb{R}^n} (v.grad\ f(x))\zeta(x)dx$$

pour tout $\zeta \in C_0^\infty(\mathbb{R}^n)$, implique :
$D_v f(x) = v.grad\ f(x)$ presque partout, sachant que $D_v f(x)$ et $v.grad\ f(x)$ sont localement intégrables.
Pour cela on rappelle que si u $\in \mathbb{L}^1_{loc}(\mathbb{R}^n)$ la distribution associée à u, T_u est définie par :

$$\forall \phi \in C_0^\infty(\mathbb{R}^n) \ \ < T_u, \phi >= \int_{\mathbb{R}^n} u(x)\phi(x)dx$$

On a alors le théorème suivant :

Théorème 10
Soit $u \in \mathbb{L}^1_{loc}(\mathbb{R}^n)$.
Alors la distribution associée à u est nulle, en tant que distribution, si et seulement si u est nulle presque partout.

Démonstration :
Soit $\phi \in C_0^\infty(\mathbb{R}^n)$ et soit K le support de ϕ. Alors

$$| < T_u, \phi > | \leq (\sup_K |\phi|) \int_K |u(x)|dx$$

Si u est nulle presque partout, il est clair que T_u est nulle.
Réciproquement, supposons T_u est identiquement nulle.
Soit K un compact inclus dans $\mathbb{R}^n$. On va montrer que u, restreinte à K, est nulle.
Soient $\delta > 0$, et soit :

$$K_{j\delta} = \{x \in \mathbb{R}^n / d(x, K) \leq j\delta\}$$

On note u_0 la restriction de u à K_δ et $\widetilde{u_0}$ la fonction égale à u_0 sur K_δ et nulle ailleurs.
Soit $(j_\epsilon)_{\epsilon>0}$ une famille régularisante, ie une famille de fonctions telles que :

$$\begin{aligned}
&i. j_\epsilon \in C_0^\infty(\mathbb{R}^n)\\
&ii. j_\epsilon \geq 0\\
&iii. \textstyle\int_{\mathbb{R}^n} j_\epsilon(x)dx \ = \ 1\\
&iv.\ supp(j_\epsilon) \ = \ B_f(0,\epsilon)
\end{aligned}$$

Alors on a :

$$u_\epsilon \ = \ \widetilde{u_0} \star j_\epsilon \in C_0^\infty(\mathbb{R}^n)$$

Soit $0 < \epsilon < \delta$. Pour $x \in K$:

$$u_\epsilon(x) = \int_{\mathbb{R}^n} \widetilde{u_0}(y)\ j_\epsilon(x-y)dy = \int_{|x-y|\leq\epsilon} \widetilde{u_0}(y)\ j_\epsilon(x-y)dy$$

Donc $y \in K_\epsilon \subseteq K_\delta$. Donc :

$$u_\epsilon(x) \;=\; \int_{K_\delta} \widetilde{u_0}(y)\; j_\epsilon(x-y)dy = \int_{K_\delta} u_0(y)\; j_\epsilon(x-y)dy = \int_{K_\delta} u(y)\; j_\epsilon(x-y)dy = 0$$

Donc si $x \in K$, $u_\epsilon(x) = 0$.
$\widetilde{u_0} \in \mathbb{L}^1(\mathbb{R}^n)$ à support compact. Donc u_ϵ converge dans $\mathbb{L}^1(\mathbb{R}^n)$ vers $\widetilde{u_0}$. En particulier, la restriction de u_ϵ à K converge vers la restriction de $\widetilde{u_0}$ dans $\mathbb{L}^1(K)$. Donc :

$$\lim_{\epsilon \to 0} \int_K |\widetilde{u_0}(x) - u_\epsilon(x)|dx = \lim_{\epsilon \to 0} \int_K |\widetilde{u_0}(x)|dx = \lim_{\epsilon \to 0} \int_K |u(x)|dx = 0$$

D'où :

$$\int_K |u(x)|dx = 0$$

Donc la restriction de u à K est nulle presque partout.

$\square$

Références

[1] Leçons d'analyse fonctionnelle
Frederic RIESZ et Béla SZ.-NAGY

[2] Measure theory and fine properties of functions
Lawrence C. EVANS et Ronald F. GARIEPY

[3] Cours de mathématiques spéciales. Topologie et éléments d'analyse.
E. RAMIS, C. DESCHAMPS et J. ODOUX

[4] Intégration.
André GRAMAIN

[5] Courbes et dimension fractale
Claude TRICOT

$$\boxed{\textbf{Fonctions continues non dérivables}}$$