ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Section de mathématiques (SMA)
Printemps 2016

Gradient, dérivées directionnelles, différentielle

Version du 1er mars 2016
Responsable : B. Buffoni

Table des matières

1 Introduction

Dans ce chapitre, sauf mention explicite du contraire, les fonctions seront définies sur des sous-ensembles *ouverts* de $\mathbb{R}^n$ ($n \geq 1$). Ainsi, lorsque nous écrirons par exemple $f : E \to \mathbb{R}$ ou $\mathbf{f} : E \to \mathbb{R}^m$ ($m \geq 1$), il sera sous-entendu que E est ou sous-ensemble ouvert de $\mathbb{R}^n$.

Il est usuel d'écrire les éléments de $\mathbb{R}^n$ sous la forme de n-uples :

$$\mathbf{x} = \vec{x} = \underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n;$$

voir par exemple le chapitre 11 du livre de Douchet-Zwahlen [DZ]. Dans ce chapitre, nous n'utiliserons la notation $(x_1, \ldots, x_n)$ que lorsque les composantes de $\mathbf{x}$ apparaissent dans l'argument d'une fonction $f : E \to \mathbb{R}$ ou $\mathbf{f} : E \to \mathbb{R}^m$. Ainsi nous écrirons par exemple $f(x_1, \ldots, x_n) \in \mathbb{R}$. Dans les autres cas, sauf mention explicite du contraire, nous utiliserons la notation sous forme de vecteur colonne :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

A l'aide de la transposée, ceci s'écrit aussi

$$\mathbf{x} = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix}^T$$

ou encore

$$\mathbf{x} = (x_1\,,\ \ldots\ ,\, x_n)^T\,.$$

La somme de deux vecteurs dans $\mathbb{R}^n$ devient

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ y_1 + y_n \end{pmatrix},$$

le produit d'un nombre réel λ avec un vecteur $\mathbf{x} \in \mathbb{R}^n$ devient

$$\lambda\mathbf{x} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

et le produit scalaire usuel dans $\mathbb{R}^n$ devient

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^{n} x_j y_j\,.$$

Le j-ème vecteur de la base canonique de $\mathbb{R}^n$

$$\mathbf{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

a toutes ses composantes nulles, sauf celle sur la j-ème ligne, qui vaut 1.

Pour des entiers $n, m \geq 1$, nous dirons dans ces notes qu'une application $L : \mathbb{R}^n \to \mathbb{R}^m$ est *linéaire* si

$$L(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha L(\mathbf{x}) + \beta L(\mathbf{y})$$

pour tous $\alpha, \beta \in \mathbb{R}$ et $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. En choisissant $\alpha = \beta = 0$ et $\mathbf{x} = \mathbf{y} = 0$, ceci implique en particulier que l'image de $\mathbf{0} \in \mathbb{R}^n$ par L est $\mathbf{0} \in \mathbb{R}^m$:

$$L(\mathbf{0}) = \mathbf{0}.$$

L'espace vectoriel de toutes les applications linéaires $L : \mathbb{R}^n \to \mathbb{R}^m$ sera noté par $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

2 Dérivées partielles, gradient

Soit un entier $n \geq 1$. Soient encore un sous-ensemble (ouvert) $E \subset \mathbb{R}^n$ et une fonction $f : E \to \mathbb{R}$ de la variable $\mathbf{x} = (x_1, \ldots, x_n) \in E$.

Pour $\mathbf{a} \in E$ et $1 \leq j \leq n$ fixés, on considère la fonction réelle g définie par

$$g(s) = f(a_1, a_2, \ldots, a_{j-1}, s, a_{j+1}, \ldots, a_n),$$

son domaine de définition étant le sous-ensemble ouvert de $\mathbb{R}$ donné par

$$D(g) = \Big\{ s \in \mathbb{R} : \, (a_1, a_2, \ldots, a_{j-1}, s, a_{j+1}, \ldots, a_n)^T \in E \Big\}.$$

Rappel et notations
Si g est dérivable en a_j, on dit que *la j-ème dérivée partielle* de f existe en $\mathbf{a}$ et est égale à $g'(a_j)$. Elle est notée

$$\frac{\partial f}{\partial x_j}(\mathbf{a}) \quad \text{ou} \quad D_j f(\mathbf{a}).$$

Comme

$$g'(a_j) = \lim_{t \to 0} \frac{g(a_j + t) - g(a_j)}{t},$$

on voit que

$$\boxed{\frac{\partial f}{\partial x_j}(\mathbf{a}) = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{e}_j) - f(\mathbf{a})}{t}.}$$

Si toutes les dérivées partielles $\frac{\partial f}{\partial x_1}(\mathbf{a}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{a})$ existent en $\mathbf{a} \in E$, le vecteur

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

est appelé le *gradient* de f en $\mathbf{a}$, que l'on note $\nabla f(\mathbf{a})$ ou $\mathbf{grad}\, f(\mathbf{a})$:

$$\boxed{\nabla f(\mathbf{a}) = \mathbf{grad}\, f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} D_1 f(\mathbf{a}) \\ \cdot \\ \cdot \\ \cdot \\ D_n f(\mathbf{a}) \end{pmatrix}.}$$

3 Dérivées directionnelles

Soient $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$, $\mathbf{a} \in E$ et $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. La ligne droite passant par $\mathbf{a}$ et admettant le vecteur directeur $\mathbf{v}$ est paramétrée comme suit :

$$\{\mathbf{a} + t\mathbf{v} : t \in \mathbb{R}\}.$$

Soit la fonction réelle g définie par $g(t) = f(\mathbf{a}+t\mathbf{v})$, son domaine de définition étant $D(g) = \{t \in \mathbb{R} : \mathbf{a} + t\mathbf{v} \in E\}$. C'est un sous-ensemble ouvert de $\mathbb{R}$ qui contient 0.

Si g est dérivable en $t = 0$, on dit que f est *dérivable en* $\mathbf{a}$ *suivant le vecteur* $\mathbf{v}$. De plus $g'(0)$ est appelé *la dérivée de* f *en* $\mathbf{a}$ *suivant* $\mathbf{v}$ ou encore *la dérivée directionnelle de* f *en* $\mathbf{a}$ *suivant* $\mathbf{v}$, que l'on note

$$Df(\mathbf{a}, \mathbf{v}) \quad \text{ou} \quad \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}).$$

Ainsi

$$Df(\mathbf{a}, \mathbf{v}) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}) = \lim_{t\to 0} \frac{g(t) - g(0)}{t} = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t}.$$

Si $\mathbf{v} = \mathbf{e}_j$ et f est dérivable en $\mathbf{a}$ suivant le vecteur $\mathbf{e}_j$, observons que $Df(\mathbf{a}, \mathbf{e}_j) = \frac{\partial f}{\partial x_j}(\mathbf{a})$.

Ainsi, si f est dérivable en $\mathbf{a}$ suivant tous les vecteurs $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, alors toutes les dérivées partielles $\frac{\partial f}{\partial x_j}(\mathbf{a})$, $j = 1, \ldots, n$, sont bien définies.

La réciproque n'est pas vraie en général. Voir l'exemple 1 du §5.

Pour $\lambda \in \mathbb{R} \backslash \{0\}$,

$$\begin{aligned} Df(\mathbf{a}, \lambda \mathbf{v}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\lambda \mathbf{v}) - f(\mathbf{a})}{t} &= \lim_{s \to 0} \frac{f(\mathbf{a} + s\mathbf{v}) - f(\mathbf{a})}{s/\lambda} \\ &= \lambda \lim_{s \to 0} \frac{f(\mathbf{a} + s\mathbf{v}) - f(\mathbf{a})}{s} = \lambda Df(\mathbf{a}, \mathbf{v}), \end{aligned}$$

où $s = \lambda t$.

Ainsi, si $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ et f est dérivable en $\mathbf{a}$ suivant $\mathbf{v}$, f est aussi dérivable en $\mathbf{a}$ suivant $\lambda \mathbf{v}$ pour tout $\lambda \in \mathbb{R} \backslash \{0\}$ et

$$\boxed{Df(\mathbf{a}, \lambda \mathbf{v}) = \lambda Df(\mathbf{a}, \mathbf{v}).}$$

Il suffit donc de considérer les dérivées suivant les vecteurs unitaires. Lorsque $||\mathbf{v}|| = 1$, on dit aussi *dérivée dans la direction* $\mathbf{v}$.

4 Fonction dérivable, différentielle

La définition 4.1 ci-dessous de *différentielle* est une extension de la notion de *différentiabilité* d'une fonction réelle d'une variable réelle (voir le §5.1.3 dans [DZ]). Rappelons qu'une fonction $f : E \to \mathbb{R}$ définie sur un sous-ensemble ouvert $E \subset \mathbb{R}$ est dite *différentiable* en $a \in E$ s'il existe $\ell \in \mathbb{R}$ et une fonction $r : E \to \mathbb{R}$ vérifiant les deux propriétés suivantes :

$$\forall x \in E \quad f(x) = f(a) + \ell(x - a) + r(x)$$

et

$$\lim_{x \to a} \frac{r(x)}{x - a} = 0.$$

Ceci est le cas exactement lorsque f est dérivable en a et

$$\ell = f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \in \mathbb{R}$$

(§5.1.4 dans [DZ]). De plus, dans ce cas, le graphique de f admet une droite tangente passant par $(a, f(a))$ d'équation cartésienne

$$y = f(a) + f'(a)(x - a)$$

(§5.1.7 dans [DZ]). La définition 4.1 ci-dessous de différentielle pour les fonctions $f : E \to \mathbb{R}$ avec $E \subset \mathbb{R}^n$ admet également une interprétation géométrique lorsque $n = 2$, qui est expliquée au §4.2. La droite tangente y est remplacée par un *plan tangent*.

Il résulte du dernier résultat fondamental du §4.3 une manière simple de vérifier que $f : E \to \mathbb{R}$ est différentiable en tout point de l'ensemble (ouvert) $E \subset \mathbb{R}^n$: il suffit que les n dérivées partielles de f existent en tout point de E et que chacune soit continue sur E. Cette condition *suffisante* est souvent utilisée dans la pratique.

4.1 Définition : différentielle

Soient $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$ et $\mathbf{a} \in E$.

On dit que f est *dérivable au point* $\mathbf{a}$ ou *différentiable au point* $\mathbf{a}$ s'il existe une application *linéaire* $L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ et une fonction $r : E \to \mathbb{R}$ telles que

$$\forall \mathbf{x} \in E \quad f(\mathbf{x}) = f(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + r(\mathbf{x})$$

et

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{r(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = 0,$$

où $L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) \in \mathbb{R}$ est la valeur de l'application $L_{\mathbf{a}}$ évaluée en $\mathbf{x} - \mathbf{a}$.

Il y a au plus une application linéaire $L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ vérifiant cette condition. Lorsque f est différentiable en $\mathbf{a}$, $L_{\mathbf{a}}$ est appelé *la différentielle* de f au point $\mathbf{a}$. On utilise la notation suivante pour $L_{\mathbf{a}}$:

$$L_{\mathbf{a}} = df(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}),$$

où $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est l'espace vectoriel des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$.

Remarques.
• Soit une application linéaire $L : \mathbb{R}^n \to \mathbb{R}$. Définissons

$$\boldsymbol{\ell} = \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_n \end{pmatrix} \in \mathbb{R}^n$$

par

$$\boldsymbol{\ell} = \begin{pmatrix} L(\mathbf{e}_1) \\ \vdots \\ L(\mathbf{e}_n) \end{pmatrix}.$$

Alors

$$\forall \mathbf{v} \in \mathbb{R}^n \quad L(\mathbf{v}) = < \boldsymbol{\ell}, \mathbf{v} >,$$

où $< \cdot, \cdot >$ est le produit scalaire usuel dans $\mathbb{R}^n$. En effet, pour tout $\mathbf{v} \in \mathbb{R}^n$,

$$L(\mathbf{v}) = L\left(\sum_{j=1}^{n} v_j \mathbf{e}_j\right) = \sum_{j=1}^{n} v_j L(\mathbf{e}_j) = \sum_{j=1}^{n} v_j \ell_j =< \boldsymbol{\ell}, \mathbf{v} > .$$

• Lorsque $n = 1$, f admet une différentielle en a exactement lorsque la dérivée $f'(a) \in \mathbb{R}$ existe.
• Les adjectifs “différentiable” et “dérivable” seront toujours considérés comme synonymes.
• Si f est dérivable en tout $\mathbf{a} \in E$, alors $df : E \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est une fonction à valeurs dans l'espace vectoriel $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$. Nous dirons que f est dérivable ou différentiable *sur* E et que $df : E \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est la (fonction) différentielle de f.

4.2 Interprétation géométrique lorsque $n = 2$

Soient un sous-ensemble (ouvert) $E \subset \mathbb{R}^2$ et une fonction $f : E \to \mathbb{R}$ dérivable en $\mathbf{a} \in E$ de différentielle $L_{\mathbf{a}} = df(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^2, \mathbb{R})$. Soit $\boldsymbol{\ell}$ tel que

$$\forall \mathbf{v} \in \mathbb{R}^2 \quad L_{\mathbf{a}}(\mathbf{v}) =< \boldsymbol{\ell}, \mathbf{v} > .$$

L'équation

$$x_3 = f(\mathbf{a}) + \langle \boldsymbol{\ell}, \mathbf{x} - \mathbf{a} \rangle = f(a_1, a_2) + \ell_1(x_1 - a_1) + \ell_2(x_2 - a_2), \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2,$$

décrit un plan dans $\mathbb{R}^3$ passant par le point

$$\begin{pmatrix} a_1 \\ a_2 \\ f(a_1, a_2) \end{pmatrix} .$$

La direction normale à ce plan est donnée par le vecteur

$$\begin{pmatrix} -\ell_1 \\ -\ell_2 \\ 1 \end{pmatrix}$$

car

$$x_3 = f(\mathbf{a}) + \langle \boldsymbol{\ell}, \mathbf{x} - \mathbf{a} \rangle \Leftrightarrow \left\langle \begin{pmatrix} -\ell_1 \\ -\ell_2 \\ 1 \end{pmatrix} , \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \\ x_3 - f(a_1, a_2) \end{pmatrix} \right\rangle = 0.$$

Si f est dérivable en $\mathbf{a} \in E \subset \mathbb{R}^2$, $L_{\mathbf{a}} = df(\mathbf{a})$ et $\boldsymbol{\ell} \in \mathbb{R}^2$ est tel que

$$\forall \mathbf{v} \in \mathbb{R}^2 \quad L_{\mathbf{a}}(\mathbf{v}) =< \boldsymbol{\ell}, \mathbf{v} >,$$

alors l'équation

$$x_3 = f(\mathbf{a}) + \langle \boldsymbol{\ell}, \mathbf{x} - \mathbf{a} \rangle\,, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

décrit le plan dans $\mathbb{R}^3$ tangent au graphique de f en

$$\begin{pmatrix} a_1 \\ a_2 \\ f(a_1, a_2) \end{pmatrix} \in \mathbb{R}^3.$$

La normale au graphique de f en $(a_1, a_2, f(a_1, a_2))^T$ est la droite paramétrée comme suit :

$$\left\{ \begin{pmatrix} a_1 \\ a_2 \\ f(a_1, a_2) \end{pmatrix} + t \begin{pmatrix} -\ell_1 \\ -\ell_2 \\ 1 \end{pmatrix} \;:\; t \in \mathbb{R} \right\}.$$

4.3 Résultats fondamentaux

Théorème. Soient $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$ et $\mathbf{a} \in E$ avec f dérivable en $\mathbf{a}$ de différentielle $L_{\mathbf{a}} = df(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. Alors

1. f est continue en $\mathbf{a}$.
2. Pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, f est dérivable en $\mathbf{a}$ suivant le vecteur $\mathbf{v}$ et

$$Df(\mathbf{a}, \mathbf{v}) = L_{\mathbf{a}}(\mathbf{v}).$$

En particulier

$$\frac{\partial f}{\partial x_j}(\mathbf{a}) = L_{\mathbf{a}}(\mathbf{e}_j)$$

pour tout $j \in \{1, \ldots, n\}$. Autrement dit, le gradient $\nabla f(\mathbf{a})$ existe et

$$\nabla f(\mathbf{a}) = \boldsymbol{\ell},$$

où $\ell_j = L_{\mathbf{a}}(\mathbf{e}_j)$ pour tout $j \in \{1, \ldots, n\}$.

3.

Pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$,

$$Df(\mathbf{a}, \mathbf{v}) = L_{\mathbf{a}}(\mathbf{v}) = < \nabla f(\mathbf{a}), \mathbf{v} > .$$

4.

Pour tout $\mathbf{v} \in \mathbb{R}^n$ tel que $||\mathbf{v}|| = 1$,

$$Df(\mathbf{a}, \mathbf{v}) \leq ||\nabla f(\mathbf{a})||$$

et, si $\nabla f(\mathbf{a}) \neq \mathbf{0}$,

$$Df\left(\mathbf{a}, \frac{1}{||\nabla f(\mathbf{a})||} \nabla f(\mathbf{a})\right) = ||\nabla f(\mathbf{a})||.$$

Ainsi $\frac{1}{||\nabla f(\mathbf{a})||} \nabla f(\mathbf{a}) \in \mathbb{R}^n$ donne la *direction de la plus grande pente* de f en $\mathbf{a}$ et $||\nabla f(\mathbf{a})||$ est la valeur de la plus grande pente.

Remarques.
• Il découle de la relation $L_{\mathbf{a}}(\mathbf{v}) = Df(\mathbf{a}, \mathbf{v})$ pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ que la différentielle $L_{\mathbf{a}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est unique si elle existe, comme ceci a déjà été mentionné.
• La seconde partie du théorème assure que si f est dérivable en $\mathbf{a}$, alors f est dérivable en $\mathbf{a}$ suivant tout vecteur $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. La réciproque n'est pas vraie en général ; voir l'exemple 2 du §5.

Démonstration. 1. Vérifions que $\lim\limits_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. On a

$$\forall \mathbf{x} \in E \quad f(\mathbf{x}) = f(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + r(\mathbf{x})$$

avec

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{r(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = 0.$$

Il en résulte $\lim\limits_{\mathbf{x} \to \mathbf{a}} r(\mathbf{x}) = 0$ et

$$\begin{aligned} \lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) &= \lim_{\mathbf{x} \to \mathbf{a}} \Big(f(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + r(\mathbf{x}) \Big) \\ &= f(\mathbf{a}) + \lim_{\mathbf{x} \to \mathbf{a}} L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + \lim_{\mathbf{x} \to \mathbf{a}} r(\mathbf{x}) = f(\mathbf{a}). \end{aligned}$$

2. Fixons $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ et posons $g(t) = f(\mathbf{a} + t\mathbf{v})$. Alors $D(g)$ est un sous-ensemble ouvert de $\mathbb{R}$ qui contient 0 et il faut démontrer que g est dérivable

en $t = 0$. Or en effet

$$
\begin{aligned}
g'(0) &= \lim_{t\to 0} \frac{g(t)-g(0)}{t} = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{v})-f(\mathbf{a})}{t} \\
&= \lim_{t\to 0} \frac{f(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{a}+t\mathbf{v}-\mathbf{a}) + r(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t} \\
&= \lim_{t\to 0} \frac{L_{\mathbf{a}}(t\mathbf{v}) + r(\mathbf{a}+t\mathbf{v})}{t} = L_{\mathbf{a}}(\mathbf{v}) + \lim_{t\to 0} \frac{r(\mathbf{a}+t\mathbf{v})}{t} \\
&= L_{\mathbf{a}}(\mathbf{v}) + \lim_{t\to 0} \frac{r(\mathbf{a}+t\mathbf{v})}{||\mathbf{a}+t\mathbf{v}-\mathbf{a}||} \, \frac{||\mathbf{a}+t\mathbf{v}-\mathbf{a}||}{t} \\
&= L_{\mathbf{a}}(\mathbf{v}) + \lim_{t\to 0} \frac{r(\mathbf{a}+t\mathbf{v})}{||\mathbf{a}+t\mathbf{v}-\mathbf{a}||} \, \frac{|t|\,||\mathbf{v}||}{t} = L_{\mathbf{a}}(\mathbf{v}),
\end{aligned}
$$

ce qui montre que $Df(\mathbf{a},\mathbf{v}) = g'(0) = L_{\mathbf{a}}(\mathbf{v})$.

En choisissant en particulier $\mathbf{v} = \mathbf{e}_j$, on obtient

$$
\frac{\partial f}{\partial x_j}(\mathbf{a}) = Df(\mathbf{a},\mathbf{e}_j) = L_{\mathbf{a}}(\mathbf{e}_j) = \ell_j \, .
$$

D'où

$$
\boldsymbol{\ell} = \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix} = \nabla f(\mathbf{a}).
$$

3. Pour tout $\mathbf{v} \in \mathbb{R}^n$,

$$
L_{\mathbf{a}}(\mathbf{v}) = L_{\mathbf{a}}\left(\sum_{j=1}^{n} v_j \mathbf{e}_j\right) = \sum_{j=1}^{n} v_j L_{\mathbf{a}}(\mathbf{e}_j) = \sum_{j=1}^{n} v_j \ell_j = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(\mathbf{a}) = < \nabla f(\mathbf{a}), \mathbf{v} > .
$$

4. Pour tout $\mathbf{v} \in \mathbb{R}^n$ tel que $||\mathbf{v}|| = 1$,

$$
D(\mathbf{a},\mathbf{v}) = < \nabla f(\mathbf{a}), \mathbf{v} > \leq ||\nabla f(\mathbf{a})|| \cdot ||\mathbf{v}|| = ||\nabla f(\mathbf{a})||
$$

et, si $\nabla f(\mathbf{a}) \neq \mathbf{0}$,

$$
D\left(\mathbf{a}, \frac{1}{||\nabla f(\mathbf{a})||} \nabla f(\mathbf{a})\right) = \left\langle \nabla f(\mathbf{a}), \frac{1}{||\nabla f(\mathbf{a})||} \nabla f(\mathbf{a}) \right\rangle = ||\nabla f(\mathbf{a})||.
$$

$\square$

Le théorème suivant donne une méthode pour vérifier qu'une fonction est dérivable en un point.

Théorème. Soient $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$ et $\mathbf{a} \in E$.

Supposons qu'il existe $\delta > 0$ tel que $B(\mathbf{a}, \delta) \subset E$ et, pour tout $j \in \{1, \ldots, n\}$, la dérivée partielle $\frac{\partial f}{\partial x_j}$ existe sur $B(\mathbf{a}, \delta)$ et est continue en $\mathbf{a}$. Alors f est dérivable en $\mathbf{a}$.

Remarque. Ceci améliore la conclusion du §13.1.6 dans [DZ].

Démonstration. Soient $\mathbf{x} \in \mathbb{R}^n$ tel que $||\mathbf{x} - \mathbf{a}|| < \delta$. Alors

$$\begin{aligned}
& f(\mathbf{x}) - f(\mathbf{a}) = f(x_1, x_2, \ldots, x_n) - f(a_1, a_2, \ldots, a_n) \\
& = f(x_1, a_2, \ldots, a_n) - f(a_1, a_2, \ldots, a_n) \\
& \quad + f(x_1, x_2, a_3, \ldots, a_n) - f(x_1, a_2, \ldots, a_n) \\
& \quad + \ldots \\
& \quad + f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{n-1}, a_n) \\
& = \frac{\partial f}{\partial x_1}\Big(a_1 + \theta_1(x_1 - a_1), a_2, \ldots, a_n\Big)(x_1 - a_1) \\
& + \frac{\partial f}{\partial x_2}\Big(x_1, a_2 + \theta_2(x_2 - a_2), a_3, \ldots, a_n\Big)(x_2 - a_2) \\
& \quad + \ldots \\
& + \frac{\partial f}{\partial x_n}\Big(x_1, x_2, \ldots, x_{n-1}, a_n + \theta_n(x_n - a_n)\Big)(x_n - a_n),
\end{aligned}$$

où $\theta_j \in]0, 1[$ pour $j = 1, \ldots, n$ est donné par la formule des accroissements finis pour une fonction réelle d'une variable réelle. Ainsi

$$\begin{aligned}
& f(\mathbf{x}) - f(\mathbf{a}) - \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle \\
& = \frac{\partial f}{\partial x_1}\Big(a_1 + \theta_1(x_1 - a_1), a_2, \ldots, a_n\Big)(x_1 - a_1) - \frac{\partial f}{\partial x_1}(a_1, a_2, \ldots, a_n)(x_1 - a_1) \\
& + \frac{\partial f}{\partial x_2}\Big(x_1, a_2 + \theta_2(x_2 - a_2), a_3, \ldots, a_n\Big)(x_2 - a_2) - \frac{\partial f}{\partial x_2}(a_1, a_2, a_3, \ldots, a_n)(x_2 - a_2) \\
& + \ldots \\
& + \frac{\partial f}{\partial x_n}\Big(x_1, x_2, \ldots, x_{n-1}, a_n + \theta_n(x_n - a_n)\Big)(x_n - a_n) - \frac{\partial f}{\partial x_n}(a_1, a_2, \ldots, a_n)(x_n - a_n)
\end{aligned}$$

et donc

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle}{||\mathbf{x} - \mathbf{a}||} = 0$$

grâce à la continuité des dérivées partielles en $\mathbf{a}$. Ainsi f est dérivable en $\mathbf{a}$ et $df(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est l'application linéaire

$$\mathbf{v} \to \langle \nabla f(\mathbf{a}), \mathbf{v} \rangle .$$

□

5 Exemples

5.1 Exemple 1

Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x_1, x_2) = \begin{cases} 0 & \text{si } \mathbf{x} = \mathbf{0}, \\ \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{si } \mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}. \end{cases}$$

Remarquons d'abord que f n'est pas continue en $\mathbf{0}$ car

$$\lim_{t \to 0} f(t, t) = \frac{1}{2} \neq 0 = f(\mathbf{0})$$

(voir le §12.2.9 dans [DZ]). Donc f n'est pas dérivable en $\mathbf{0}$. Son gradient $\nabla f(x_1, x_2)$ est bien défini en tout $\mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$ et

$$\nabla f(x_1, x_2) = \begin{pmatrix} \dfrac{x_2(x_1^2 + x_2^2) - x_1 x_2 2 x_1}{(x_1^2 + x_2^2)^2} \\ \\ \dfrac{x_1(x_1^2 + x_2^2) - x_1 x_2 2 x_2}{(x_1^2 + x_2^2)^2} \end{pmatrix} = \begin{pmatrix} \dfrac{x_2(-x_1^2 + x_2^2)}{(x_1^2 + x_2^2)^2} \\ \\ \dfrac{x_1(x_1^2 - x_2^2)}{(x_1^2 + x_2^2)^2} \end{pmatrix}.$$

Puisque $\frac{\partial f}{\partial x_1}$ et $\frac{\partial f}{\partial x_2}$ sont continues en tout point de l'ouvert $\mathbb{R}^2 \backslash \{\mathbf{0}\}$, f est dérivable sur $\mathbb{R}^2 \backslash \{\mathbf{0}\}$ et la dérivée directionnelle de f en $\mathbf{a} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$ suivant $\mathbf{v} \neq \mathbf{0}$ peut se calculer à l'aide du gradient :

$$D(\mathbf{a}, \mathbf{v}) = < \nabla f(\mathbf{a}), \mathbf{v} > .$$

La différentielle de f en $\mathbf{a} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$ est l'application linéaire

$$\mathbf{v} \to < \nabla f(a_1, a_2), \mathbf{v} > = \frac{a_2(-a_1^2 + a_2^2) v_1}{(a_1^2 + a_2^2)^2} + \frac{a_1(a_1^2 - a_2^2) v_2}{(a_1^2 + a_2^2)^2}$$

($\mathbf{v} \in \mathbb{R}^2$). En $\mathbf{a} = \mathbf{0}$, les dérivées partielles existent :

$$\frac{\partial f}{\partial x_1}(\mathbf{0}) = \lim_{t \to 0} \frac{f(t,0) - f(0,0)}{t} = \lim_{t \to 0} \frac{1}{t} \left(\frac{t \cdot 0}{t^2 + 0^2} - 0 \right) = 0$$

et de même

$$\frac{\partial f}{\partial x_2}(\mathbf{0}) = 0,$$

d'où $\nabla f(\mathbf{0}) = \mathbf{0}$. Par contre observons que les limites $\lim_{t \to 0} \dfrac{\partial f}{\partial x_1}(0, t)$ et $\lim_{t \to 0} \dfrac{\partial f}{\partial x_2}(t, 0)$ n'existent pas.

Finalement, pour $\mathbf{v} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$,

$$D(\mathbf{0}, \mathbf{v}) = \lim_{t \to 0} \frac{f(tv_1, tv_2) - f(\mathbf{0})}{t} = \lim_{t \to 0} \frac{v_1 v_2}{t(v_1^2 + v_2^2)}$$

est bien défini dans $\mathbb{R}$ si et seulement si $v_1 = 0$ ou $v_2 = 0$.

Remarque. En plus des notations $f(x_1, x_2)$, $\frac{\partial f}{\partial x_1}$ et $\frac{\partial f}{\partial x_2}$, les notations $f(x, y)$, $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ sont souvent utilisées. Voir par exemple les §13.1.3 et 13.1.4 dans [DZ].

5.2 Exemple 2

Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x_1, x_2) = \begin{cases} 0 & \text{si } \ \mathbf{x} = \mathbf{0}, \\ \dfrac{x_1 x_2^2}{x_1^2 + x_2^4} & \text{si } \ \mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}. \end{cases}$$

Remarquons que f n'est pas continue en $\mathbf{0}$ car

$$\lim_{t \to 0} f(t^2, t) = \lim_{t \to 0} \frac{t^2 \cdot t^2}{t^4 + t^4} = \frac{1}{2} \neq 0 = f(\mathbf{0}).$$

Donc f n'est pas dérivable en $\mathbf{0}$. Son gradient $\nabla f(x_1, x_2)$ est bien défini en tout $\mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$ et

$$\nabla f(x_1, x_2) = \begin{pmatrix} \dfrac{x_2^2(x_1^2 + x_2^4) - 2x_1^2 x_2^2}{(x_1^2 + x_2^4)^2} \\ \\ \dfrac{2x_1 x_2(x_1^2 + x_2^4) - 4x_1 x_2^5}{(x_1^2 + x_2^4)^2} \end{pmatrix} = \begin{pmatrix} \dfrac{x_2^2(-x_1^2 + x_2^4)}{(x_1^2 + x_2^4)^2} \\ \\ \dfrac{2x_1 x_2(x_1^2 - x_2^4)}{(x_1^2 + x_2^4)^2} \end{pmatrix}.$$

Puisque $\frac{\partial f}{\partial x_1}$ et $\frac{\partial f}{\partial x_2}$ sont continues en tout point de l'ouvert $\mathbb{R}^2 \backslash \{\mathbf{0}\}$, f est dérivable sur $\mathbb{R}^2 \backslash \{\mathbf{0}\}$. En $\mathbf{a} = \mathbf{0}$, les dérivées partielles existent :

$$\frac{\partial f}{\partial x_1}(\mathbf{0}) = \lim_{t \to 0} \frac{f(t, 0) - f(0, 0)}{t} = \lim_{t \to 0} \frac{1}{t} \left(\frac{t \cdot 0^2}{t^2 + 0^4} - 0 \right) = 0$$

et de même

$$\frac{\partial f}{\partial x_2}(\mathbf{0}) = 0,$$

d'où $\nabla f(\mathbf{0}) = \mathbf{0}$. Par contre observons que

$$\lim_{t \to 0} \frac{\partial f}{\partial x_1}(t, t) = \lim_{t \to 0} \frac{t^2(-t^2 + t^4)}{(t^2 + t^4)^2} = -1 \neq 0 = \frac{\partial f}{\partial x_1}(\mathbf{0}).$$

Plus généralement, pour tout $\mathbf{v} \in \mathbb{R}^2 \backslash \{0\}$, la dérivée de f en $\mathbf{0}$ suivant $\mathbf{v}$ existe :

$$D(\mathbf{0}, \mathbf{v}) = \lim_{t \to 0} \frac{1}{t} \left(\frac{(tv_1)(tv_2)^2}{(tv_1)^2 + (tv_2)^4} - 0 \right) = \lim_{t \to 0} \frac{v_1 v_2^2}{(v_1)^2 + t^2 v_2^4} = \frac{v_2^2}{v_1} \in \mathbb{R}$$

si $v_1 \neq 0$ et, si $v_1 = 0$,

$$D(\mathbf{0}, \mathbf{v}) = \lim_{t \to 0} \frac{1}{t} \left(\frac{(t \cdot 0)(tv_2)^2}{(t \cdot 0)^2 + (tv_2)^4} - 0 \right) = 0.$$

En posant $D(\mathbf{0}, \mathbf{0}) = 0$, l'application $\mathbf{v} \to D(\mathbf{0}, \mathbf{v})$ n'est clairement pas linéaire en $\mathbf{v} \in \mathbb{R}^2$ dans cet exemple. De plus, pour $\mathbf{v} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$, la relation $D(\mathbf{0}, \mathbf{v}) =< \nabla f(\mathbf{0}), \mathbf{v} >$ n'est valable que si $v_1 = 0$ ou $v_2 = 0$.

6 Fonctions à valeurs dans $\mathbb{R}^m$

Soit $f : E \to \mathbb{R}$ avec $E \subset \mathbb{R}^n$ ouvert. Si le gradient de f est défini en tout $\mathbf{a} \in E$, la fonction gradient

$$\mathbf{a} \to \nabla f(\mathbf{a})$$

est un exemple d'une fonction définie sur E et à valeurs dans $\mathbb{R}^n$. En particulier, dans le premier exemple ci-dessus, le gradient

$$\nabla f(x_1, x_2) = \begin{cases} (0,0)^T & \text{si } \mathbf{x} = \mathbf{0}, \\ \left(\dfrac{x_2(-x_1^2 + x_2^2)}{(x_1^2 + x_2^2)^2}, \dfrac{x_1(x_1^2 - x_2^2)}{(x_1^2 + x_2^2)^2} \right)^T & \text{si } \mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}, \end{cases}$$

est une fonction à valeurs dans $\mathbb{R}^2$.

Plus généralement, le but est maintenant d'étudier les fonctions à valeurs dans $\mathbb{R}^m$, où $m \geq 1$ est un entier quelconque, dans le même esprit que les considérations précédentes sur les fonctions à valeurs dans $\mathbb{R}$.

Soient des entiers $m, n \geq 1$, $E \subset \mathbb{R}^n$ (ouvert) et une fonction quelconque $\mathbf{f} : E \to \mathbb{R}^m$. Ecrivons, pour tout $\mathbf{x} \in E$,

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^m,$$

où $f_i : E \to \mathbb{R}$ pour $i = 1, \ldots, m$.

6.1 Définitions : dérivée partielle, dérivée suivant un vecteur

On dit que $\mathbf{f}$ admet une *dérivée partielle au point* $\mathbf{a} \in E$ si chacune des fonctions f_i admet une dérivée partielle au point $\mathbf{a}$. Une dérivée partielle de $\mathbf{f}$ est notée par un vecteur colonne :

$$\frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_j}(\mathbf{a}) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\mathbf{a}) \end{pmatrix}.$$

Pour un vecteur $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, la *dérivée (directionnelle) de* $\mathbf{f}$ *en* $\mathbf{a}$ *suivant* $\mathbf{v}$ est le vecteur colonne noté par $D\mathbf{f}(\mathbf{a}, \mathbf{v})$ ou $\frac{\partial \mathbf{f}}{\partial \mathbf{v}}(\mathbf{a})$ et défini par

$$D\mathbf{f}(\mathbf{a}, \mathbf{v}) = \frac{\partial \mathbf{f}}{\partial \mathbf{v}}(\mathbf{a}) = \begin{pmatrix} Df_1(\mathbf{a}, \mathbf{v}) \\ Df_2(\mathbf{a}, \mathbf{v}) \\ \vdots \\ Df_m(\mathbf{a}, \mathbf{v}) \end{pmatrix}.$$

Lorsque $||\mathbf{v}|| = 1$, on dit aussi *dérivée dans la direction* v.

Remarquons que

$$D\mathbf{f}(\mathbf{a}, \mathbf{v}) = \begin{pmatrix} \sum_{j=1}^n \frac{\partial f_1}{\partial x_j}(\mathbf{a}) v_j \\ \vdots \\ \sum_{j=1}^n \frac{\partial f_m}{\partial x_j}(\mathbf{a}) v_j \end{pmatrix} = \sum_{j=1}^n v_j \frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{a})$$

pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{0\}$.

6.2 Définition : différentielle

La notion de différentielle se généralise facilement. Soient $E \subset \mathbb{R}^n$, $\mathbf{f} : E \to \mathbb{R}^m$ et $\mathbf{a} \in E$.

On dit que $\mathbf{f}$ est *dérivable au point* $\mathbf{a}$ ou *différentiable au point* $\mathbf{a}$ s'il existe une application *linéaire* $L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m$ et une fonction $\mathbf{r} : E \to \mathbb{R}^m$ telles que

$$\forall \mathbf{x} \in E \quad \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + \mathbf{r}(\mathbf{x})$$

et

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{1}{||\mathbf{x} - \mathbf{a}||} \mathbf{r}(\mathbf{x}) = 0.$$

Il y a au plus une application linéaire $L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m$ vérifiant cette condition.

Lorsque $\mathbf{f}$ est différentiable en $\mathbf{a}$, cette application linéaire $L_{\mathbf{a}}$ est appelée *la différentielle* de $\mathbf{f}$ au point $\mathbf{a}$. On utilise les notations suivantes pour $L_{\mathbf{a}}$:

$$L_{\mathbf{a}} = d\mathbf{f}(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m).$$

6.3 Proposition

Soit $f : E \to \mathbb{R}^m$ et $\mathbf{a} \in E \subset \mathbb{R}^n$. Alors $\mathbf{f}$ est dérivable en $\mathbf{a}$ si et seulement si f_i est dérivable en $\mathbf{a}$ pour tout $i \in \{1, \ldots, m\}$. Dans ce cas, en notant

$$L_{\mathbf{a}} = d\mathbf{f}(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \quad \text{et} \quad L_{i,\mathbf{a}} = df_i(\mathbf{a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$$

pour $i \in \{1, \ldots, m\}$, on a

$$L_{\mathbf{a}}(\mathbf{v}) = \begin{pmatrix} L_{1,\mathbf{a}}(\mathbf{v}) \\ \vdots \\ L_{m,\mathbf{a}}(\mathbf{v}) \end{pmatrix} = \begin{pmatrix} Df_1(\mathbf{a}, \mathbf{v}) \\ \vdots \\ Df_m(\mathbf{a}, \mathbf{v}) \end{pmatrix} = D\mathbf{f}(\mathbf{a}, \mathbf{v})$$

pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{0\}$.

Remarque. Il découle de cette proposition que la différentielle $L_{\mathbf{a}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est unique si elle existe, comme ceci a déjà été mentionné.

6.4 Matrice et déterminant de Jacobi

Comme $\mathbb{R}^n$ et $\mathbb{R}^m$ sont munis des bases canoniques, à toute application linéaire $L : \mathbb{R}^n \to \mathbb{R}^m$ est associée une matrice à m lignes et n colonnes. Plus précisément, la j-ième colonne de cette matrice est constituée des m composantes du vecteur $L(\mathbf{e}_j)$ relativement à la base canonique de $\mathbb{R}^m$.

En particulier, si $f : E \to \mathbb{R}^m$ est dérivable en $\mathbf{a} \in E \subset \mathbb{R}^n$, on peut considérer la matrice associée à $L_{\mathbf{a}} = df(\mathbf{a})$. Sa j-ième colonne est

$$L_{\mathbf{a}}(\mathbf{e}_j) = \begin{pmatrix} L_{1,\mathbf{a}}(\mathbf{e}_j) \\ \vdots \\ L_{m,\mathbf{a}}(\mathbf{e}_j) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(\mathbf{a}) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\mathbf{a}) \end{pmatrix},$$

c'est-à-dire, la j-ième dérivée partielle de $\mathbf{f}$ en $\mathbf{a}$.

Ceci conduit à la définition suivante.

La *matrice de Jacobi* ou *matrice Jacobienne* de $\mathbf{f}$ (ou de $f_1, \ldots, f_m$) au point $\mathbf{a}$ est la matrice notée

$$D\mathbf{f}(\mathbf{a}) \quad \text{ou} \quad J_{\mathbf{f}}(\mathbf{a})$$

définie par

$$
\boxed{D\mathbf{f}(\mathbf{a}) = J_{\mathbf{f}}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}.}
$$

Elle a m lignes et n colonnes. Sa i-ème ligne est

$$
\begin{pmatrix} \frac{\partial f_i}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_i}{\partial x_n}(\mathbf{a}) \end{pmatrix},
$$

qui est la transposée du vecteur colonne $\nabla f_i(\mathbf{a})$:

$$
\begin{pmatrix} \frac{\partial f_i}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_i}{\partial x_n}(\mathbf{a}) \end{pmatrix} = \nabla f_i(\mathbf{a})^T
$$

$(1 \leq i \leq m)$. Si $\mathbf{f}$ est dérivable en $\mathbf{a}$, observons que, pour tout $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$,

$$
\boxed{D\mathbf{f}(\mathbf{a}, \mathbf{v}) = \begin{pmatrix} Df_1(\mathbf{a}, \mathbf{v}) \\ Df_2(\mathbf{a}, \mathbf{v}) \\ \vdots \\ Df_m(\mathbf{a}, \mathbf{v}) \end{pmatrix} = \begin{pmatrix} \langle \nabla f_1(\mathbf{a}), \mathbf{v} \rangle \\ \langle \nabla f_2(\mathbf{a}), \mathbf{v} \rangle \\ \vdots \\ \langle \nabla f_m(\mathbf{a}), \mathbf{v} \rangle \end{pmatrix} = D\mathbf{f}(\mathbf{a}) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.}
$$

Lorsque $m = n$, le déterminant

$$
\det D\mathbf{f}(\mathbf{a}) = |D\mathbf{f}(\mathbf{a})| = \begin{vmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{a}) & \frac{\partial f_n}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_n}{\partial x_n}(\mathbf{a}) \end{vmatrix}
$$

est appelé le *déterminant de Jacobi* ou le *Jacobien* de f (ou de $f_1, \dots, f_n$) au point $\mathbf{a}$. Il est noté aussi $\dfrac{D(f_1, \dots, f_n)}{D(x_1, \dots, x_n)}(\mathbf{a})$:

$$
\boxed{\frac{D(f_1, \dots, f_n)}{D(x_1, \dots, x_n)}(\mathbf{a}) = \det D\mathbf{f}(\mathbf{a}) = \begin{vmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{a}) & \frac{\partial f_n}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_n}{\partial x_n}(\mathbf{a}) \end{vmatrix}.}
$$

6.5 Fonction composée et matrices de Jacobi

On notera par $M_{m,n}(\mathbb{R})$ les matrices à coefficients réels possédant m lignes et n colonnes.

Enoncé.

Soient un sous-ensemble ouvert $A \subset \mathbb{R}^n$, une fonction $\mathbf{g} : A \to \mathbb{R}^p$, un sous-ensemble ouvert $B \subset \mathbb{R}^p$ tel que $\mathbf{g}(A) \subset B$ et une fonction $\mathbf{f} : B \to \mathbb{R}^q$. La fonction composée $\mathbf{f} \circ \mathbf{g} : A \to \mathbb{R}^q$ est donc bien définie.

Soient encore $\mathbf{a} \in A$ et $\mathbf{b} = \mathbf{g}(\mathbf{a}) \in B$ tels que $\mathbf{g}$ est dérivable en $\mathbf{a}$ et $\mathbf{f}$ est dérivable en $\mathbf{b}$.

Alors $\mathbf{f} \circ \mathbf{g}$ est dérivable en $\mathbf{a}$,

$$\boxed{d(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = d\mathbf{f}(\mathbf{b}) \circ d\mathbf{g}(\mathbf{a}), \quad \text{où} \quad \mathbf{b} = \mathbf{g}(\mathbf{a}),}$$

et la matrice de Jacobi $D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) \in M_{q,n}(\mathbb{R})$ est le produit matriciel des matrices de Jacobi $D\mathbf{f}(\mathbf{b}) \in M_{q,p}(\mathbb{R})$ et $D\mathbf{g}(\mathbf{a}) \in M_{p,n}(\mathbb{R})$:

$$\boxed{D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{b}) \cdot D\mathbf{g}(\mathbf{a}), \quad \text{où} \quad \mathbf{b} = \mathbf{g}(\mathbf{a}).}$$

Si de plus $n = p = q$, alors on obtient la relation suivante pour les déterminants de Jacobi :

$$\boxed{|D(\mathbf{f} \circ \mathbf{g})(\mathbf{a})| = |D\mathbf{f}(\mathbf{b})| \cdot |D\mathbf{g}(\mathbf{a})| \quad \text{où} \quad \mathbf{b} = \mathbf{g}(\mathbf{a}).}$$

Démonstration. Nous noterons par $\mathbf{y}$ l'argument variable de la fonction $\mathbf{g}$ et par $\mathbf{x}$ celui de la fonction $\mathbf{f}$. Ecrivons $L_{\mathbf{a}} = d\mathbf{g}(\mathbf{a})$ et $L_{\mathbf{b}} = d\mathbf{f}(\mathbf{b})$. Les applications $L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^p$ et $L_{\mathbf{b}} : \mathbb{R}^p \to \mathbb{R}^q$ étant linéaires, leur composée $L_{\mathbf{b}} \circ L_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^q$ est aussi une application linéaire. Vérifions que

$$\forall \mathbf{y} \in A \quad \mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{f}(\mathbf{g}(\mathbf{a})) + L_{\mathbf{b}}(L_{\mathbf{a}}(\mathbf{y} - \mathbf{a})) + \mathbf{r}(\mathbf{y})$$

et

$$\lim_{\mathbf{y} \to \mathbf{a}} \frac{1}{||\mathbf{y} - \mathbf{a}||} \mathbf{r}(\mathbf{y}) = 0$$

pour une certaine fonction $\mathbf{r} : A \to \mathbb{R}^q$.

Par hypothèse, il existe des fonctions $r_{\mathbf{g}} : A \to \mathbb{R}^p$ et $r_{\mathbf{f}} : B \to \mathbb{R}^q$ telles que

$$\forall \mathbf{y} \in A \quad \mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{y} - \mathbf{a}) + \mathbf{r}_{\mathbf{g}}(\mathbf{y}),$$

$$\lim_{\mathbf{y} \to \mathbf{a}} \frac{1}{||\mathbf{y} - \mathbf{a}||} \mathbf{r}_{\mathbf{g}}(\mathbf{y}) = 0$$

$$\forall \mathbf{x} \in B \quad \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{b}) + L_{\mathbf{b}}(\mathbf{x} - \mathbf{b}) + \mathbf{r}_{\mathbf{f}}(\mathbf{x})$$

et

$$\lim_{\mathbf{x} \to \mathbf{b}} \frac{1}{||\mathbf{x} - \mathbf{b}||} \mathbf{r}_{\mathbf{f}}(\mathbf{x}) = 0.$$

On a pour tout $\mathbf{y} \in A$

$$
\begin{aligned}
\mathbf{f}(\mathbf{g}(\mathbf{y})) &= \mathbf{f}\Big(\mathbf{g}(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{y}-\mathbf{a}) + \mathbf{r}_{\mathbf{g}}(\mathbf{y})\Big)\\
&= \mathbf{f}(\mathbf{b}) + L_{\mathbf{b}}\Big(L_{\mathbf{a}}(\mathbf{y}-\mathbf{a}) + \mathbf{r}_{\mathbf{g}}(\mathbf{y})\Big) + \mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y}))\\
&= \mathbf{f}(\mathbf{g}(\mathbf{a})) + L_{\mathbf{b}}(L_{\mathbf{a}}(\mathbf{y}-\mathbf{a})) + L_{\mathbf{b}}(\mathbf{r}_{\mathbf{g}}(\mathbf{y})) + \mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y})).
\end{aligned}
$$

En posant

$$\mathbf{r}(\mathbf{y}) = L_{\mathbf{b}}(\mathbf{r}_{\mathbf{g}}(\mathbf{y})) + \mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y})),$$

on obtient bien

$$\lim_{\mathbf{y}\to\mathbf{a}} \frac{1}{||\mathbf{y}-\mathbf{a}||}\mathbf{r}(\mathbf{y}) = 0.$$

En effet, pour tout $\mathbf{y} \in A\backslash\{\mathbf{a}\}$,

$$
\begin{aligned}
||L_{\mathbf{a}}(\mathbf{y}-\mathbf{a})|| &= \left\|\begin{pmatrix} <\nabla g_1(\mathbf{a}), \mathbf{y}-\mathbf{a}> \\ \vdots \\ <\nabla g_p(\mathbf{a}), \mathbf{y}-\mathbf{a}> \end{pmatrix}\right\| = \left(\sum_{i=1}^{p}(<\nabla g_i(\mathbf{a}), \mathbf{y}-\mathbf{a}>)^2\right)^{1/2}\\
&\overset{\text{Cauchy-Schwarz}}{\leq} \left(\sum_{i=1}^{p}||\nabla g_i(\mathbf{a})||^2||\mathbf{y}-\mathbf{a}||^2\right)^{1/2} = \left(\sum_{i=1}^{p}||\nabla g_i(\mathbf{a})||^2\right)^{1/2}||\mathbf{y}-\mathbf{a}||,
\end{aligned}
$$

$$
\begin{aligned}
\frac{||L_{\mathbf{b}}(\mathbf{r}_{\mathbf{g}}(\mathbf{y}))||}{||\mathbf{y}-\mathbf{a}||} &= \frac{1}{||\mathbf{y}-\mathbf{a}||}\left\|\begin{pmatrix} <\nabla f_1(\mathbf{b}), \mathbf{r}_{\mathbf{g}}(\mathbf{y})> \\ \vdots \\ <\nabla f_q(\mathbf{b}), \mathbf{r}_{\mathbf{g}}(\mathbf{y})> \end{pmatrix}\right\|\\
&\leq \frac{1}{||\mathbf{y}-\mathbf{a}||}\left(\sum_{i=1}^{q}||\nabla f_i(\mathbf{b})||^2||\mathbf{r}_{\mathbf{g}}(\mathbf{y})||^2\right)^{1/2} = \left(\sum_{i=1}^{q}||\nabla f_i(\mathbf{b})||^2\right)^{1/2}\frac{||\mathbf{r}_{\mathbf{g}}(\mathbf{y})||}{||\mathbf{y}-\mathbf{a}||}
\end{aligned}
$$

et

$$
\begin{aligned}
\frac{||\mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y}))||}{||\mathbf{y}-\mathbf{a}||} &= \begin{cases} 0 & \text{si } \ \mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{a}) = \mathbf{b},\\ \dfrac{||\mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y}))||}{||\mathbf{g}(\mathbf{y})-\mathbf{b}||}\,\dfrac{||\mathbf{g}(\mathbf{y})-\mathbf{g}(\mathbf{a})||}{||\mathbf{y}-\mathbf{a}||} & \text{si } \ \mathbf{g}(\mathbf{y}) \neq \mathbf{b}, \end{cases}\\
&= \begin{cases} 0 & \text{si } \ \mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{a}) = \mathbf{b},\\ \dfrac{||\mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y}))||}{||\mathbf{g}(\mathbf{y})-\mathbf{b}||}\,\dfrac{||L_{\mathbf{a}}(\mathbf{y}-\mathbf{a})+\mathbf{r}_{\mathbf{g}}(\mathbf{y})||}{||\mathbf{y}-\mathbf{a}||} & \text{si } \ \mathbf{g}(\mathbf{y}) \neq \mathbf{b}, \end{cases}\\
&\leq \begin{cases} 0 & \text{si } \ \mathbf{g}(\mathbf{y}) = \mathbf{g}(\mathbf{a}) = \mathbf{b},\\ \dfrac{||\mathbf{r}_{\mathbf{f}}(\mathbf{g}(\mathbf{y}))||}{||\mathbf{g}(\mathbf{y})-\mathbf{b}||}\left(\left(\displaystyle\sum_{i=1}^{p}||\nabla g_i(\mathbf{a})||^2\right)^{1/2} + \dfrac{||\mathbf{r}_{\mathbf{g}}(\mathbf{y})||}{||\mathbf{y}-\mathbf{a}||}\right) & \text{si } \ \mathbf{g}(\mathbf{y}) \neq \mathbf{b}. \end{cases}
\end{aligned}
$$

Ceci prouve que $\mathbf{f}\circ\mathbf{g}$ est dérivable en $\mathbf{a}$ et $d(\mathbf{f}\circ\mathbf{g})(\mathbf{a}) = L_{\mathbf{b}}\circ L_{\mathbf{a}}$. La théorie générale des matrices et déterminants assure finalement que

$$D(\mathbf{f}\circ\mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{b})\cdot D\mathbf{g}(\mathbf{a})$$

(produit matriciel) et, lorsque $n = p = q$,

$$|D(\mathbf{f} \circ \mathbf{g})(\mathbf{a})| = |D\mathbf{f}(\mathbf{b})| \cdot |D\mathbf{g}(\mathbf{a})|$$

(produit de deux déterminants).

□

7 Bibliographie

• S. Benzoni-Gavage, Calcul différentiel et équations différentielles - 2e éd. - Cours et exercices corrigés, Dunod, Paris, 2014.

• J. Douchet, Analyse, Recueil d'exercices et aide-mémoire, vol. 2, PPUR, 2003.

• J. Douchet et B. Zwahlen, Calcul différentiel et intégral, Fonctions réelles d'une ou de plusieurs variables réelles, PPUR, 2006.

• C. A. Stuart, Analyse I et II pour ingénieurs, semestre 2, deuxième édition, EPFL.

• J. Stubbe, Analyse II pour ingénieurs, 2011, EPFL.

• P. Wittwer, Analyse II, notes manuscrites, EPFL, printemps 2013 et 2015.