UNIVERSITE DE GENEVE — Mathématiques pour Informaticiens
Section de Mathématiques — Le 13 Avril 2018

Petit Formulaire de Calcul Différentiel

1. *Définition générale de la différentielle d'une application.*
On dit que $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ est différentiable en $a \in U$ pour U ouvert de $\mathbb{R}^n$ si il existe une application linéaire notée $f'(a) : \mathbb{R}^n \to \mathbb{R}^m$ (ou $Df(a)$) telle que pour tout $h \to 0$,
$$f(a+h) = f(a) + f'(a)(h) + o(h).$$
$f'(a)(h)$ (ou $Df(a)(h)$) est la différentielle au *point* a et dans la *direction* h.
Remarque : en 1D, on avait la dérivée à gauche et la dérivée à droite. Quand on travaille sur $\mathbb{R}^n$, on peut dériver dans une infinité de directions.
• Il est bien souvent plus simple de garder la direction et de ne pas "simplifier par h" comme on le faisait en 1D.
• On peut revenir à cette définition dès que l'on est sur des espaces de matrices ou que l'on étudie des applications multilinéaires, quadratique, etc... Par exemple, $f : E \to F$ et E ou F est $\mathcal{M}_n(\mathbb{R})$, ou si $f(x) = x^T Ax + b^T x + c$, il ne faut pas calculer la jacobienne, c'est (beaucoup) plus compliqué.
• Par exemple, si $f(A) = A^p$ où A est une matrice, on écrit
$$f(A+H) = (A+H)^p = A^p + \sum_{k=0}^{p} A^k H A^{p-k} + \text{termes contenant au moins deux } H.$$
Alors on peut identifier chaque morceau : $f(A) = A^p$, $f'(A)(H) = \sum_{k=0}^{p} A^k H A^{p-k}$ et le reste qui est un $o(H)$. On observe qu'on ne peut pas écrire très simplement $f'(A)(H)$ sous la forme $f'(A).H$ (car A est une matrice et H aussi).

2. *Calculer la jacobienne d'une application.*
Si $f : \mathbb{R}^n \to \mathbb{R}^m$, la jacobienne $f'(x)$ ou $J_f(x)$ au point x est le tableau rangeant les dérivées partielles de f évaluées en x. Dans une **CO**lonne, on range une **CO**ordonnée et cela donne
$$f'(x) = J_f(x) = \left(\frac{\partial f_i}{\partial x_j}(x) \right)_{i,j}.$$
On identifie ici la différentielle en tant qu'application linéaire $f'(x)$ avec sa matrice jacobienne $J_f(x)$. Le gradient est défini comme $\nabla f(x) = J_f(x)^T$. L'utilisation de la jacobienne est préférable lorsque chaque composante de f est composée de fonctions simples dont on sait calculer les dérivées partielles.
• Par exemple, si $f(x,y,z) = \begin{pmatrix} xy \\ x^2 \cos(z) \end{pmatrix}$ alors on évite d'écrire $f(x+h_1, y+h_2, z+h_3) = f(x,y,z) + ...$ On écrit les dérivées partielles une à une et on

les met dans le tableau. Pour calculer $\frac{\partial f_1}{\partial x}(x,y,z)$, on considère y et z comme des constantes et on fait la dérivée usuelle de $f_1(x,y,z) = xy$ par rapport à x. Ainsi $\frac{\partial f_1}{\partial x}(x,y,z) = y$. On obtient

$$J_f(x) = \begin{pmatrix} y & x & 0 \\ 2x\cos(z) & 0 & -x^2\sin(z) \end{pmatrix}.$$

3. *Compatibilité entre les définitions.*
   On a
   $$f'(x)(h) = Df(x)(h) = J_f(x).h = \sum_i \frac{\partial f}{\partial x_i}(x)h_i = \nabla f(x)^T.h.$$
   Remarque : si toutes les dérivées partielles existent et sont continues, alors la fonction est différentiable. L'existence des dérivées partielles n'entraîne pas l'existence de la différentielle.

4. *Composition de fonctions.*
   En 1D, on a la formule
   $$(f \circ g)'(x) = f'(g(x))g'(x).$$
   La formule en dimension supérieure est globalement la même. On peut l'écrire différemment selon les notations choisies :
   $$(f \circ g)'(x)(h) = f'(g(x))(g'(x)(h)),$$
   $$D(f \circ g)(x)(h) = Df(g(x))(Dg(x)(h)),$$
   $$J_{f\circ g}(x) = J_f(g(x))J_g(x).$$
   Il est pratique pour ne pas se tromper de donner des noms différents aux variables de f : par exemple, $f(y_1, ..., y_n)$ et $g(x_1, ..., x_m)$, alors
   $$J_{f\circ g}(x) = \left(\sum_{k=1}^{n} \frac{\partial f_i}{\partial y_k}(g(x)) \frac{\partial g_k}{\partial x_j}(x) \right)_{i,j}.$$

5. *Développements de Taylor.*
   Pour $f : \mathbb{R} \to \mathbb{R}$, le développement de Taylor à l'ordre N est, si f est $(N+1)$-fois continument dérivable,
   $$f(a+h) = \sum_{k=0}^{N} \frac{f^{(k)}(a)}{k!} h^k + \mathcal{O}(h^{N+1}).$$
   Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ est $(N+1)$-fois continument différentiable, h devient la direction de la différentielle :
   $$f(a+h) = \sum_{k=0}^{N} \frac{1}{k!} f^{(k)}(a)(h, h, ..., h) + \mathcal{O}(h^{N+1}).$$

Si $f : \mathbb{R}^n \to \mathbb{R}$, on a en particulier à l'ordre 2,

$$f(a+h) = f(a) + J_f(a)h + h^T H_f(a)h + o(h^2).$$

On peut réécrire le développement avec les dérivées partielles comme

$$\begin{aligned} f(a+h) = f(a) \\ &+ \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a) h_k \\ &+ \frac{1}{2!} \sum_{k=1}^{n} \sum_{p=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_p}(a) h_k h_p \\ &+ \frac{1}{3!} \sum_{k=1}^{n} \sum_{p=1}^{n} \sum_{q=1}^{n} \frac{\partial^3 f}{\partial x_k \partial x_p \partial x_q}(a) h_k h_p h_q \\ &+ \ldots \end{aligned}$$

6. *Quelques erreurs à éviter.*
• Il y a une différence entre $\frac{\partial f(xy,x^2)}{\partial x}$ et $\frac{\partial f}{\partial x}(xy, x^2)$. La première est la dérivée partielle de $f \circ g$ par rapport à x et où $g(x,y) = (xy, x^2)$, alors que $\frac{\partial f}{\partial x}(xy, x^2)$ représente la dérivée partielle de f par rapport à x évaluée en $g(x,y)$.
• La différentielle de A^p n'est pas pA. De même, la différentielle de exp en $A \in \mathcal{M}_n(\mathbb{R})$ dans la direction H n'est pas $\exp(A)H$ en général.