Université Claude Bernard–Lyon I
Agrégation interne de Mathématiques : Oral 1
Année 2012–2013

Calcul différentiel

Soient m et n deux entiers naturels non nuls, U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}^m$ une fonction. On va étudier f au voisinage d'un point fixé $x = (x_1, \ldots, x_n) \in U$.
On choisit des normes sur $\mathbb{R}^m$ et $\mathbb{R}^n$, qu'on notera toutes deux $||\cdot||$. On s'autorisera à en changer selon le moment, ce qui n'aura pas de conséquence car toutes les normes sont équivalentes. Par exemple, si f est définie au voisinage de 0, la propriété que $f(h)$ est négligeable devant $||h||$ lorsque h tend vers 0 ne dépend pas de la norme choisie ; on écrira alors : $f(h) = o(h)$.
Dans les exemples, on prendra, sauf mention expresse du contraire, $m = 1$ (fonction à valeurs réelles), $n = 1$ si le texte est en bleu et $n = 2$ si le texte est en noir. Une phrase en gris indique une idée générale qu'il faut souvent formaliser.

1° Continuité

a) Définition

La définition de la continuité de f en $x \in U$ est :

$$(n=1) \quad \forall \varepsilon > 0,\ \exists \alpha > 0,\ \forall x' \in U, \quad |x' - x| \leqslant \alpha \implies |f(x') - f(x)| \leqslant \varepsilon;$$
$$(n \text{ quelconque}) \quad \forall \varepsilon > 0,\ \exists \alpha > 0,\ \forall x' \in U, \quad ||x' - x|| \leqslant \alpha \implies ||f(x') - f(x)|| \leqslant \varepsilon.$$

Par exemple, les applications « coordonnées » $e_j^* : (x_1, \ldots, x_n) \mapsto x_j$ sont continues car elles sont, au choix, linéaires (et on est en dimension finie) ou 1-lipschitziennes ($|x_j - x'_j| \leqslant ||x - x'||_\infty$).
Voici un critère de continuité pour une fonction vectorielle (tautologique si $n = 1$) : f est continue si et seulement si ses composantes, qui sont des fonctions réelles, le sont. Par composantes, on entend les fonctions f_i telles que $f(x) = (f_1(x), \ldots, f_m(x)) \in \mathbb{R}^m$ pour $x \in U$. On le voit facilement avec la norme infinie sur $\mathbb{R}^m$.
En pratique, pour montrer la continuité, on se repose sur la démarche suivante : on prouve une fois pour toute la continuité des fonctions les plus habituelles ($(x_1, \ldots, x_n) \mapsto x_j$ pour $j \in \{1, \ldots, n\}$, cos, ln, etc.) ; on prouve la continuité de la somme, du produit, du quotient de deux fonctions continues (pour le quotient, en supposant que le dénominateur n'est pas nul) ; on prouve la continuité de la composée. Pour les cas les plus graves, on repart avec ε, α. Pour nier la continuité en x, on exhibe deux suites ou courbes qui tendent vers x dont les images par f ont des limites différentes.

b) Applications partielles, « applications radiales »

Les applications partielles associées à f sont les suivantes. On fixe $j \in \{1, \ldots, n\}$ et on note

$$U_j = \{t \in \mathbb{R},\ (x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n) \in U\}.$$

Cela permet de définir la j^{e} application partielle :

$$f^{[j]} : U_j \to \mathbb{R}^m, \quad t \mapsto f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n).$$

L'idée est de restreindre f aux axes du repère translaté dont x est l'origine. On se convainc grâce à la définition que *si f est continue en x, alors chacune des $f^{[j]}$ est continue en x_j.*
On peut généraliser cette construction à toute droite qui passe par x (au lieu des n axes de coordonnées). Pour un vecteur $v \in \mathbb{R}^n$, $v \neq 0$, on restreint f à la droite passant par x dirigée par v. Soit U_v l'ensemble des réels t pour lesquels $x + tv$ appartient à U : c'est un ouvert qui

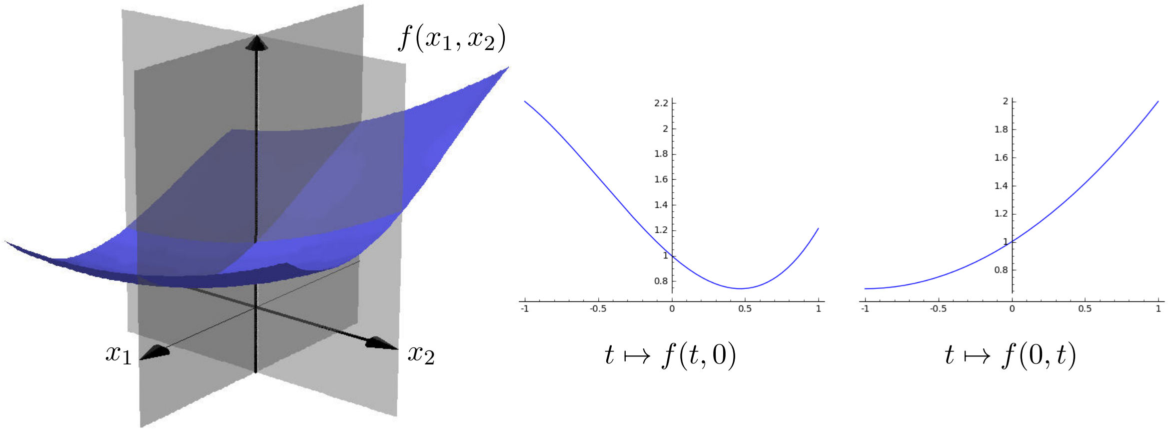

FIGURE 1 – Fonction de deux variables et deux application partielles (leurs graphes sont l'intersection du graphe de f et d'un plan vertical)

contient 0, donc un intervalle de la forme $]-\varepsilon, \varepsilon[$ pour $\varepsilon > 0$ convenable. L'application radiale [1] en x associée à v est $t \mapsto f(x + tv)$, définie sur U_v (ou sur un intervalle $]-\varepsilon, \varepsilon[$, $\varepsilon > 0$).
Une évidence : pour la continuité, seule compte la direction de v, si $v' = \alpha v$ avec $\alpha \neq 0$, alors $f(x + tv') = f\big(x + (\alpha t)v\big)$ pour tout t assez petit : les applications radiales définies par v et v' coïncident à dilatation de la variable t près.
De même que pour les applications partielles, *la continuité de f en x implique celle de toutes les applications radiales en* 0.
Malheureusement, la réciproque est fausse...

Exemple. On prend $f(x_1, x_2) = x_1/x_2$ si $x_2 \neq 0$, $f(x_1, 0) = 0$ et on se place en $x = (0, 0)$. Chaque application radiale est constante : si $v = (\cos\theta, \sin\theta)$ pour θ réel fixé, alors, si $\theta \not\equiv 0\ [\pi]$:

$$f(tv) = f(t\cos\theta, t\sin\theta) = \operatorname{cotan}\theta \quad \text{(indépendant de } t\text{)}.$$

Pourtant, f n'est pas continue ! Par exemple, elle n'est pas bornée au voisinage de 0. Ou bien, la suite définie par $p_{2k} = (2^{-k}, 0)$ et $p_{2k+1} = (2^{-k}, 2^{-k})$ converge vers 0 alors que l'image des termes pairs tend vers 0, l'image des termes impairs tend vers 1.
De même, on pourra montrer que $g(x_1, x_2) = x_1 x_2/(x_1^2 + x_2^2)$ si $(x_1, x_2) \neq 0$ ne se prolonge pas par continuité, alors que les applications radiales sont constantes sur $\mathbb{R}^*$.
En revanche, $h : 0 \mapsto 0$, $(x_1, x_2) \mapsto x_1^2 x_2/(x_1^2 + x_2^2)$ est continue en 0 car on a : $|h(x_1, x_2)| \leqslant ||(x_1, x_2)||_2$ partout.

Ça semble donc une drôle d'idée d'étudier les applications radiales ! Elles auront leur revanche.

2° Différentielle : le point de vue « abstrait »

a) Définition

On reformule la définition de la dérivabilité par un DL. Soit $a \in \mathbb{R}$.

$$\begin{aligned} f'(x) \text{ existe et vaut } a \quad &\iff \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} - a = 0 \\ &\iff \lim_{h \to 0} \frac{f(x+h) - f(x) - ah}{h} = 0 \\ &\iff f(x+h) - f(x) - ah = o(h). \end{aligned}$$

1. La terminologie n'est pas standard.

De plus, si $a' \in \mathbb{R}$ et $f'(x) = a$, alors :

$$f(x+h) - f(x) - a'h = (a - a')h + o(h) :$$

cette différence n'est négligeable devant $o(h)$ que lorsque $a' = a$. Cela implique l'unicité de la fonction affine qui approxime le mieux f au voisinage de x : c'est $x + h \mapsto f(x) + ah$. Ici, a est le coefficient directeur de l'application affine qu'est cette meilleure approximation, la différentielle étant sa partie linéaire, i.e. l'application $h \mapsto ah$.

C'est ce point de vue qui passe à plusieurs variables : la différentielle est la partie linéaire de la meilleure approximation affine de f au voisinage de x. Plus précisément, on dit que f est différentiable en x s'il existe une application linéaire $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ telle que l'on ait :

$$\forall h \in \mathbb{R}^n \text{« assez petit »}^2, \quad f(x+h) - f(x) = \varphi(h) + o(h).$$

L'idée à retenir : en cas de différentiabilité, $\Delta f = f(x+h) - f(x)$, qu'on comprend comme l'accroissement de la fonction, a une bonne approximation par une fonction linéaire de $\Delta x = (x+h) - x = h$, l'accroissement de la variable.

S'il existe une application linéaire φ, elle est unique. En effet, supposons que φ_1 et φ_2 conviennent, et soit h un vecteur assez petit, on voit en comparant les deux expressions de $f(x+h) - f(x)$: $(\varphi_1 - \varphi_2)(h) = o(h)$. Soit alors v un vecteur non nul. Pour t assez petit et $h = tv$, on a : $o(h) = o(t)$ et il vient par linéarité : $t(\varphi_1(v) - \varphi_2(v)) = o(t)$, ou : $\varphi_1(v) - \varphi_2(v) = o(1)$. En d'autres termes, comme le membre de gauche est indépendant de t : $\varphi_1(v) = \varphi_2(v)$.

Dans la situation ci-dessus, si f est différentiable en x, on note $\varphi = \mathrm{d}f_x$ et on l'appelle la différentielle de f en x.

On compare donc deux fonctions au voisinage de x : f d'une part, l'application affine $x + h \mapsto f(x) + \mathrm{d}f_x(h)$ d'autre part, dont le graphe, qui est un sous-espace affine de dimension n de $\mathbb{R}^n \times \mathbb{R}^m$ (un plan de $\mathbb{R}^3$ si $n = 2$, $m = 1$), est dit *tangent* au graphe de f.

Ainsi, par définition, si $n = 1$, f est dérivable en x SSI elle est différentiable en x et la différentielle est la multiplication par la dérivée.

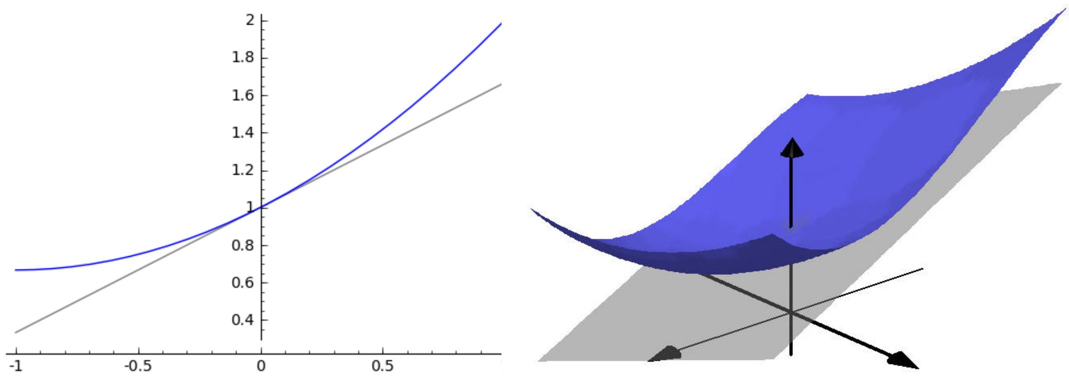

FIGURE 2 – Graphe et droite tangente, graphe et plan tangent

Exemple. Soit $f : \mathbb{R}^* \to \mathbb{R}$, $x \mapsto x^{-1}$. On a, pour $x \neq 0$ et h assez petit :

$$f(x+h) - f(x) = \frac{1}{x+h} - \frac{1}{x} = \frac{-h}{x^2\left(1 + \frac{h}{x}\right)} = -\frac{h}{x^2} + o(h).$$

Exemple. Soit $d \in \mathbb{N}^*$, $n = d^2$, $U = \mathrm{GL}_d(\mathbb{R}) \subset \mathcal{M}_d(\mathbb{R})$ et $f : U \to \mathcal{M}_d(\mathbb{R})$, $x \mapsto x^{-1}$. On montre la différentiabilité en $x = \mathrm{Id}$. Grâce à l'exemple précédent, on soupçonne que l'on a : $(\mathrm{Id} + h)^{-1} - (\mathrm{Id} - h) = o(h)$. Un calcul dans $\mathbb{R}$ s'impose :

$$\frac{1}{1+h} - (1-h) = \frac{1}{1+h} - 1 + h = \frac{-h}{1+h} + h = \frac{h^2}{1+h}.$$

On démontre alors sans peine, pour $h \in \mathcal{M}_d(\mathbb{R})$ assez petit :

$$(\mathrm{Id} + h)^{-1} - (\mathrm{Id} - h) = h^2(\mathrm{Id} + h)^{-1}.$$

Or, l'application f est continue sur U (parce que les coefficients de l'inverse d'une matrice sont des fractions rationnelles en les coefficients de la matrice), si bien que $h(\mathrm{Id} + h)^{-1}$ tend vers 0 avec h. Cela prouve que $\mathrm{d}f_{\mathrm{Id}}(h) = -h$.
Soit à présent $x \in U$ quelconque. Comme $x^{-1}h$ tend vers 0 avec h, on peut calculer :

$$\begin{aligned} f(x+h) - f(x) &= \big(x(\mathrm{Id} + x^{-1}h)\big)^{-1} - x^{-1} \\ &= (\mathrm{Id} + x^{-1}h)^{-1}x^{-1} - x^{-1} \\ &= (\mathrm{Id} - x^{-1}h + o(h))x^{-1} - x^{-1} \\ &= -x^{-1}hx^{-1} + o(h). \end{aligned}$$

Ainsi, on a : $\mathrm{d}f_x(h) = -x^{-1}hx^{-1}$ pour tout h assez petit.

b) Composition de fonctions différentiables

On se donne de plus un ouvert $V \subset \mathbb{R}^m$, $\ell \in \mathbb{N}^*$ et une fonction $g : V \to \mathbb{R}^\ell$, et on suppose que $f(U) \subset V$. On suppose de plus que f est différentiable en x et que g est différentiable en $f(x)$. Alors $g \circ f$ est différentiable en x et on a :

$$\mathrm{d}(g \circ f)_x = (\mathrm{d}g)_{f(x)} \circ (\mathrm{d}f)_x.$$

Lorsque $n = m = \ell = 1$, la différentielle de $g \circ f$ est la multiplication par $(g \circ f)'(x)$; $\mathrm{d}g_{f(x)}$ est la multiplication par $g'(f(x))$ et $\mathrm{d}f_x$ est la multiplication par $f'(x)$. Si on en croit ce qui précède, on trouve : $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$. On retrouve la formule de la dérivée de la composée habituelle.
C'est très important mais très facile. Pour $h \in \mathbb{R}^n$ assez petit, on a :

$$g \circ f(x+h) - g \circ f(x) = g\Big(f(x) + (\mathrm{d}f_x(h) + o(h))\Big) - g(f(x)).$$

Or, en dimension finie, les applications linéaires sont continues donc $(\mathrm{d}f_x(h) + o(h))$ tend vers 0 avec h, et :

$$\begin{aligned} g \circ f(x+h) - g \circ f(x) &= \mathrm{d}g_{f(x)}\big(\mathrm{d}f_x(h) + o(h)\big) + o\big(\mathrm{d}f_x(h) + o(h)\big) \\ &= \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x(h) + \mathrm{d}g_{f(x)}(o(h)) + o\big(\mathrm{d}f_x(h) + o(h)\big). \end{aligned}$$

La continuité de $\mathrm{d}f_x$ et de $\mathrm{d}g_{f(x)}$ donnent des majorations de la forme $||\mathrm{d}f_x(h)|| \leqslant C \cdot ||h||$ et $||\mathrm{d}g_{f(x)}(h')|| \leqslant C' \cdot ||h'||$, qui permettent de voir que les derniers termes sont négligeables devant $||h||$ et de conclure.

c) Dérivée selon un vecteur

Comme précédemment, on va restreindre f aux droites qui passent par x. On fixe donc $v \in \mathbb{R}^n$, $v \neq 0$ sinon ce n'est pas intéressant. On suppose que f est différentiable en x. Pour t réel assez petit, on pose : $g_v(t) = f(x + tv)$. On calcule :

$$g_v(t) = f(x + tv) = f(x) + \mathrm{d}f_x(tv) + o(tv) = f(x) + t\,\mathrm{d}f_x(v) + o(t).$$

Cela signifie que g_v est dérivable en 0 et que

$$g_v'(0) = \mathrm{d}f_x(v).$$

On définit alors en général *la dérivée de f en x dans la direction (ou le long) du vecteur v* comme cette valeur : $g_v'(0) = \mathrm{d}f_x(v)$. On vient de voir que si f est différentiable, elle admet des dérivées dans toutes les directions.

Exemple. La réciproque est fausse en général. Voici un exemple : prenons $n = 2$ et $f(x_1, x_2) = x_1^2 x_2/(x_1^2 + x_2^2)$. Pour θ réel fixé et $a > 0$, la dérivée le long de $v = (a\cos\theta, a\sin\theta)$ est :

$$g_v'(0) = a\cos^2\theta\sin\theta, \quad \text{car} \quad g_v(t) = ta\cos^2\theta\sin\theta.$$

Cela signifie que f admet des dérivées dans toutes les directions. On remarque que si v est l'un des vecteurs de la base canonique (e_1, e_2), on a : $g_{e_1}'(0) = 0 = g_{e_2}'(0)$. Ainsi, si f était différentiable, on aurait, pour $h = (h_1, h_2) \in \mathbb{R}^2$:

$$\mathrm{d}f_0(h_1, h_2) = \mathrm{d}f_0(h_1e_1 + h_2e_2) = h_1\mathrm{d}f_0(e_1) + h_2\mathrm{d}f_0(e_2) = 0,$$

ce qui contredirait, lorsque par exemple $h_1 = h_2 = 1$, l'égalité : $g_{e_1+e_2}'(0) = 1/2$.

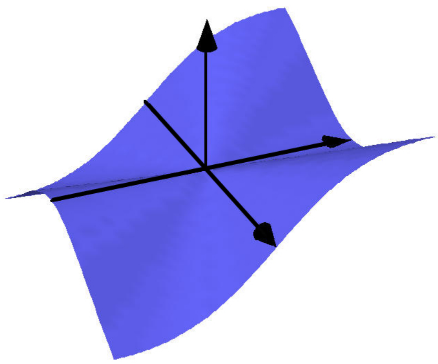

Figure 3 – Le plan qui contient l'image des axes est horizontal mais la dérivée le long de la première bissectrice n'est pas nulle

Mise en garde. J'ai trouvé ici ou là une définition de « dérivée dans la direction d'un vecteur non nul v » ce qui serait ici la dérivée le long du vecteur normalisé $v/||v||$.

Exercice. Vérifier que si $\gamma :]-\varepsilon, \varepsilon[\to \mathbb{R}^n$ est une courbe dérivable telle que $\gamma(0) = x$, alors $f \circ \gamma$ est une courbe dans $\mathbb{R}^m$ dont le vecteur tangent en 0 est $\mathrm{d}f_x(\gamma'(0))$.

3° Le point de vue « coordonnées » (fonctions à valeurs réelles)

On suppose provisoirement que $m = 1$, c'est-à-dire que f est à valeurs réelles.

a) Dérivées partielles

Pour $j \in \{1, \ldots, n\}$, on définit la j^{e} dérivée partielle de f en x comme la dérivée de f en x le long de e_j (si elle existe). Notation :

$$\frac{\partial f}{\partial x_j}(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(x + te_j).$$

C'est donc, si elle existe, la dérivée en 0 de l'application partielle

$$f^{[j]} : x_j + t \mapsto f(x_1, \ldots, x_{j-1}, x_j + t, x_{j+1}, \ldots, x_n)$$

définie sur un voisinage de x_j (dans lequel on paramètre un élément quelconque comme $x_j + t$, avec t « petit »). D'où le nom.

Attention, dans la définition de la dérivée partielle, x_j désigne deux choses différentes : dans le membre de droite, c'est le nom de la j^{e} coordonnée du système [3] ; dans le membre de gauche, c'est la j^{e} coordonnée de x, un réel fixé.

Supposons un instant que f soit différentiable en x. On a vu qu'alors, f est dérivable dans toute direction et que la dérivée directionnelle selon le vecteur v est l'image de v par $\mathrm{d}f_x$. En appliquant ceci à un vecteur de la base canonique $(e_1, \ldots, e_n)$ de $\mathbb{R}^n$, on voit que

- si f est différentiable en x, elle admet des dérivées partielles ;
- la j^{e} dérivée partielle de f est :

$$\frac{\partial f}{\partial x_j}(x) = \mathrm{d}f_x(e_j).$$

b) Critère de différentiabilité

On a vu que l'existence de dérivées directionnelles en tout point ne suffit pas à assurer la différentiabilité de f. C'est l'heure de la revanche.

Théorème. *Si f possède des dérivées partielles continues sur U, alors f est différentiable en tout point de U. De plus, f est même de classe $\mathcal{C}^1$.*

L'hypothèse signifie que toutes les fonctions $\frac{\partial f}{\partial x_j}$ sont bien définies sur U et qu'elles sont continues comme *fonctions de plusieurs variables* (ce n'est donc pas seulement l'application $f_i^{[j]}$, fonction d'une variable, qui est continue !). Le théorème est donc une réciproque, sous des hypothèses assez fortes, de la dernière remarque du paragraphe précédent.

La version $\mathcal{C}^1$ de la conclusion signifie que l'application « différentielle de f », c'est-à-dire

$$\mathrm{d}f : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}), \quad x \mapsto \mathrm{d}f_x$$

est continue comme fonction de U dans l'espace de dimension finie formé par les applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$.

c) Matrice de la différentielle

Dans $\mathbb{R}^n$, on note $(e_1, \ldots, e_n)$ la base canonique et les $(x_1, \ldots, x_n)$ coordonnées correspondantes, c'est-à-dire la base duale : chaque x_i est l'application linéaire qui, à un élément de $\mathbb{R}^n$, associe sa i^{e} coordonnée. Alors, f associe à tout vecteur-colonne à n lignes appartenant à U, un scalaire.

On suppose que f est différentiable en x : sa différentielle en x est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$, dont la matrice dans les bases $(e_1, \ldots, e_n)$ (base canonique de $\mathbb{R}^n$) et 1 (base canonique de $\mathbb{R}$) est la matrice à n colonnes et 1 ligne dont la j^{e} colonne est $\frac{\partial f}{\partial x_j}(x)$. La j^{e} colonne est l'image

3. La j^{e} coordonnée est une forme linéaire sur $\mathbb{R}^n$: c'est l'application qui à $(h_1, \ldots, h_n)$ associe h_j ; c'est quasiment une indéterminée.

par $\mathrm{d}f_x$ du j^{e} vecteur de base, e_j. D'après ce qu'on a vu, $\mathrm{d}f_x(e_j)$ est la j^{e} dérivée partielle. La matrice de $\mathrm{d}f_x$ est donc :

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Reformulons cela par un développement limité à l'ordre 1 : si f est différentiable en x, on a :

$$f(x+h) = f(x) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \cdot h_j + o(h).$$

d) Aparté : droites dans le plan

Soit a, b, c trois réels avec $v = (a, b) \neq (0, 0)$. On s'intéresse à la droite D_k du plan (coordonnées : (x_1, x_2) qui a pour équation

$$ax_1 + bx_2 + c - k = 0,$$

où k est, disons, un entier. Toutes ces droites sont parallèles entre elles et perpendiculaires à v (figure 4, à gauche).

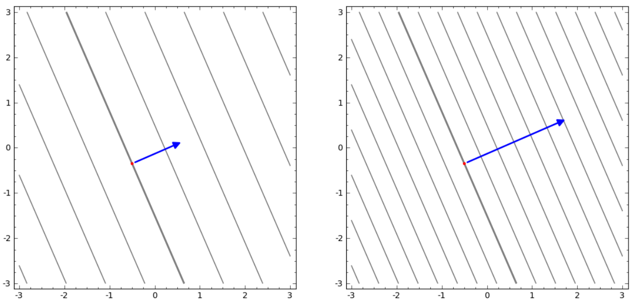

Figure 4 – Lignes de niveaux d'une fonction affine et gradient en un point

En introduisant la fonction affine

$$f : (x, y) \mapsto ax_1 + bx_2 + c,$$

on interprète les droites D_k comme les lignes de niveau k de f. Alors, la différentielle de f en un point quelconque est l'application $\mathrm{d}f_x : (h_1, h_2) \mapsto ah_1 + bh_2$, que l'on peut interpréter en termes du produit scalaire usuel : $\mathrm{d}f_x(h) = \langle v, h \rangle$. On a : $v = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$, on appellera v le *gradient* de f au point x (comme f est affine, le gradient est constant).

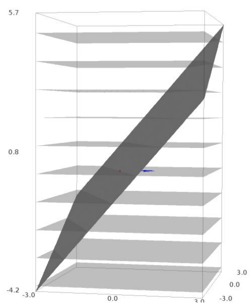

Figure 5 – Graphe d'une fonction affine (foncé) et plans horizontaux : les projections des intersections sont les lignes de niveaux

Le graphe de f est le plan gris foncé d'équation $x_3 = ax_1 + bx_2 + c$. Les lignes de niveau sont les projections dans le plan horizontal (Ox_1x_2) des intersections de ce plan et des plans horizontaux. La ligne de plus grande pente se comprend bien : partant du point rouge, dans quelle direction faut-il partir pour gagner le plus rapidement possible de l'altitude ? Réponse : dans la direction de (a, b), bien sûr ! C'est-à-dire, dans la direction du gradient de f.

À présent, on multiplie l'équation initiale par 2 et on considère les lignes de niveau ℓ entier (figure 4, à droite) :

$$2ax_1 + 2bx_2 + 2c - \ell = 0.$$

On voit que le vecteur normal (a, b) a été remplacé par $(2a, 2b)$, les lignes de niveau sont deux fois plus serrées.
Retenir : la direction du gradient indique la ligne de plus grande pente ; sa norme est d'autant plus grande que les lignes de niveau sont serrées, c'est-à-dire le profil « pentu » ou « abrupt ».

e) Gradient

Le gradient n'a de sens que pour les fonctions à valeurs réelles – c'est-à-dire lorsque $m = 1$.
Il s'agit d'étendre les considérations précédentes au cas des fonctions différentiables. L'idée, c'est que les dessins précédents sont alors ce que l'on observe si on regarde le graphe de très près, en l'identifiant au voisinage d'un point à son hyperplan tangent.
On munit $\mathbb{R}^n$ de son produit scalaire canonique, noté $\langle \cdot, \cdot \rangle$ – celui qui fait de la base canonique une base orthonormée.
Quand on a un produit scalaire, toute forme linéaire – par exemple la différentielle d'une fonction scalaire en un point – peut se représenter comme le produit scalaire par un vecteur. En effet, une forme linéaire est de la forme $\ell : (h_j) \mapsto \sum_{j=1}^n a_j h_j$, où (a_j) est fixé. On interprète ℓ comme le produit scalaire par le vecteur (a_j).
Appliquons cette remarque à la différentielle $\mathrm{d}f_x$. Pour $h = \sum_{j=1}^n h_j e_j$, on a donc :

$$\mathrm{d}f_x(h) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x) \cdot h_j = \langle \nabla f_x, h \rangle \quad \text{où} \quad \nabla f_x = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

Par construction, le gradient est orthogonal aux lignes de niveau de f. Cela signifie que si on fixe un point x sur une ligne de niveau [4] k, et si on trace une courbe sur cette ligne de niveau, alors le vecteur tangent à la courbe en x est orthogonal au gradient.
Ce n'est pas mystérieux : si $\gamma :]-\varepsilon, \varepsilon[\to N_k$ est une courbe dérivable telle que $\gamma(0) = x$, alors, puisque la fonction $g = f \circ \gamma :]-\varepsilon, \varepsilon[\to \mathbb{R}$ est constante égale à k, sa dérivée en 0 est nulle. Or, par composition, on a :

$$0 = g'(0) = \mathrm{d}f_x(\gamma'(0)) = \langle \nabla f_x, \gamma'(0) \rangle.$$

D'autre part, le gradient indique la ligne de plus grande pente. En termes imagés, si le graphe de f représente un profil de montagne, où f indique l'altitude, la direction donnée par le gradient indique la façon la plus efficace de gagner de l'altitude au voisinage de x. Si $n = 2$, on regarde, parmi les plans verticaux qui contiennent x, lequel coupe le graphe de f selon la courbe la plus pentue.
De façon un peu plus précise, fixons un cap et voyons combien on gagne en altitude en partant de x dans cette direction. Cela revient – encore une fois – à étudier la restriction de f à la droite passant par x et dirigée par le cap choisi.

4. La ligne ou surface de niveau $k \in \mathbb{R}$ est l'ensemble $N_k = f^{-1}(k)$ des points x de $\mathbb{R}^n$ tels que $f(x) = k$. C'est génériquement une courbe lorsque $n = 2$, ce qu'assure le théorème des fonctions implicites.

De façon précise, pour v vecteur de norme 1, on considère (encore !) la fonction $g_v :]-\varepsilon, \varepsilon[\to \mathbb{R}$, $t \mapsto f(x+tv)$ et on cherche pour quel v la dérivée de g_v en 0 est maximale. Puisque $g_v'(0) = \mathrm{d}f_x(v) = \langle \nabla f_x, v \rangle$, on voit que $g_v'(0)$ est maximale et égale à $||\nabla f_x||$ exactement lorsque v est positivement colinéaire au gradient ∇f_x.
Au passage, on voit que pour annuler la pente $g_v'(0)$, il faut choisir v orthogonal au gradient, c'est-à-dire tangent à la ligne/surface de niveau.

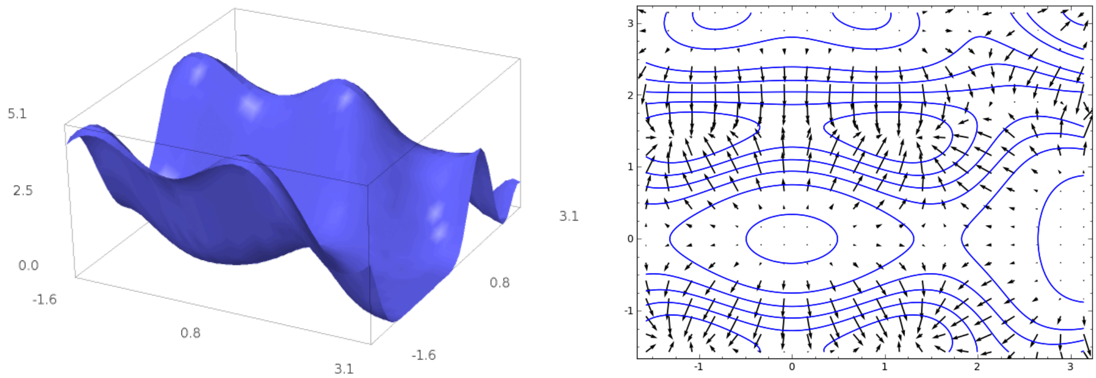

FIGURE 6 – Le graphe d'une fonction (à g.) ; lignes de niveau et gradient (à dr.). On reconnaît le minimum en $(0,0)$, le col vers $(1.6, 0)$, les deux bosses sur le bord et sur les droites $y = 1.6$...

Si on s'intéresse à l'hyperplan tangent au graphe de f en x, qu'on veut représenter dans $\mathbb{R}^{n+1}$ muni des coordonnées $(x_1, \ldots, x_n, z)$, c'est le graphe de l'approximation affine $x + h \mapsto f(x+h) + \langle \nabla f_x, h \rangle$. Un point $(X_1, \ldots, X_n, Z)$ appartient à ce plan si et seulement si on a :

$$Z = f(x) + \langle \nabla f_x, X - x \rangle = f(x) + \sum_{j=1}^{n} (X_j - x_j) \frac{\partial f}{\partial x_j}(x).$$

(Si $n = 1$, on retrouve la droite $Z = f(x) + (X - x) f'(x)$.)

Remarque (Canonicité du gradient). Si on change de base, les dérivées partielles de f ou la matrice de la différentielle changent ; mais si les bases de départ et d'arrivée sont toutes deux orthonormées, alors elles

Exercice. On trouvera cinq exercices interactifs (WIMS) ici.

4° Le point de vue « coordonnées » (fonctions à valeurs vectorielles)

a) Si m est quelconque à nouveau, on note $(x_1, \ldots, x_n)$ (resp. $(y_1, \ldots, y_m)$) le système de coordonnées de $\mathbb{R}^n$ (resp. $\mathbb{R}^m$) correspondant au repère affine standard $(O, e_1, \ldots, e_n)$ (resp. $(O, e_1', \ldots, e_m')$). Alors, f ressemble à ça :

$$\begin{array}{cccc} f : & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m, \\ & \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} & \longmapsto & \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}. \end{array}$$

On voit facilement que f est différentiable en un point x SSI chaque f_i l'est. En effet, f_i est la composée de f et d'une application linéaire, donc différentiable : la i^{e} projection. Inversement, si chaque f_i est différentiable, on peut vérifier que f l'est et que

$$\mathrm{d}f_x(h) = \begin{pmatrix} (\mathrm{d}f_1)_x(h) \\ \vdots \\ (\mathrm{d}f_m)_x(h) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \frac{\partial f_1}{\partial x_j}(x) \cdot h_j \\ \vdots \\ \sum_{j=1}^n \frac{\partial f_m}{\partial x_j}(x) \cdot h_j \end{pmatrix}.$$

En particulier, si les $\partial f_i/\partial f_j$ sont continues sur U, alors f est de classe $\mathcal{C}^1$, c'est-à-dire que l'application « différentielle de f », c'est-à-dire

$$\mathrm{d}f : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), \quad x \mapsto \mathrm{d}f_x$$

est continue comme fonction de U dans l'espace de dimension finie formé par les applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$.

b) Matrice de la différentielle

De ce qui précède, on peut déduire la matrice de $\mathrm{d}f_x$ dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$. C'est, bien sûr, la matrice $m \times n$ dont la j^{e} colonne est la colonne des coordonnées de $\mathrm{d}f_x(e_j)$. Le terme d'indice (i, j), avec $1 \leqslant i \leqslant m$ et $1 \leqslant j \leqslant n$, est :

$$\frac{\partial f_i}{\partial x_j}(x).$$

Pour ne pas confondre la matrice et sa transposée, retenir le cas d'une application linéaire $X \mapsto AX$; on peut aussi dire que les vecteurs de $\mathbb{R}^n$ et $\mathbb{R}^m$ se disposent en colonnes (pour permettre la multiplication par une matrice) et que donc, les *lignes* de la matrice de la différentielle suivent

5° Inversion locale et globale

a) Préliminaire : local et global

Parlons simplement de fonctions. Une propriété *locale* est une propriété qui concerne un voisinage d'un point, voisinage dont la taille n'est pas connue *a priori*. Pour un exemple, une limite, la continuité en un point, la dérivabilité en un point sont des hypothèses locales. La formule de Taylor-Young est une propriété locale – avec des notations censées transparentes, elle exprime que $(f(x) - T_{f,x_0,n}(x))/(x - x_0)^n$ a une limite en x_0.

A contrario, une propriété *globale* porte sur un intervalle (plus généralement un ensemble) défini à l'avance. Par exemple, l'injectivité ou la surjectivité, la continuité uniforme sont des propriétés globales ; la formule de Taylor-reste intégral(e ?) est une bien sûr formule globale.

Un bon nombre de théorèmes d'analyse sont des théorèmes de *passage du local au global*. Par exemple, les propriétés usuelles des fonctions continues sont de cette nature (valeurs intermédiaires ; une fonction continue sur un compact est bornée et atteint ses bornes ; continue sur un compact $\Rightarrow$ uniformément continue) : on sait quelque chose sur un voisinage non précisé de chaque point, on en déduit une propriété globale. Le théorème des accroissements finis aussi : sous des hypothèses locales (dérivabilité sur l'intervalle ouvert et continuité sur l'intervalle fermé), on a une hypothèse globale : $(f(b) - f(a))(b - a)$ appartient à l'image de f' ; on peut en déduire que la dérivée d'une fonction satisfait à une propriété de valeurs intermédiaires (Darboux). Idem pour la formule de Taylor-Lagrange.

b) Inversion : dimension 1 et au-delà

Il est très facile d'inverser des fonctions réelles d'une variable réelle : toute fonction continue et strictement monotone est une bijection sur son ensemble d'arrivée. Pour une fonction dérivable, cela donne lieu au critère bien connu : sous une hypothèse locale ($\pm f' > 0$ partout), on a une conclusion globale : l'injectivité de f. C'est l'ordre sur $\mathbb{R}$ qui permet de passer de l'un à l'autre (la clé, c'est la monotonie donnée par le théorème des accroissements finis).

Reformulons pour extrapoler en dimension supérieure. Pour une fonction $\mathcal{C}^1$ sur un intervalle, l'hypothèse pertinente est : $f'(x) \neq 0$ pour tout x. Pour étendre, on se souvient que $f'(x)$ est le coefficient directeur de la différentielle, c'est-à-dire que la différentielle est inversible : sous cette forme, l'hypothèse a un sens en toute dimension.

L'absence d'ordre dans $\mathbb{R}^n$ pose néanmoins un problème nouveau en dimension supérieure et empêche de passer du local au global. On peut garder en tête l'exemple des coordonnées polaires : l'application

$$f : \mathbb{R}^{+*} \times \mathbb{R} \longrightarrow \mathbb{R}^2 \setminus \{(0,0)\}, \quad (\rho, \theta) \longmapsto (\rho\cos\theta, \rho\sin\theta)$$

est de classe $\mathcal{C}^1$, sa différentielle est inversible en tout point, mais elle n'est pas globalement injective car $f(\rho, \theta + 2\pi) = f(\rho, \theta)$ pour tout (ρ, θ).

c) Énoncé local

Théorème. *Soit f définie d'un ouvert de $\mathbb{R}^n$ dans $\mathbb{R}^n$ et x^0 un point en lequel la différentielle $\mathrm{d}f_{x^0}$ est inversible. Alors, il existe un voisinage V de x^0 et un voisinage W de $f(x^0)$ tel que :*
- *la restriction de f à V est un difféomorphisme, c'est-à-dire une bijection dont la réciproque f^{-1} est de classe $\mathcal{C}^1$;*
- *pour $y \in W$, la différentielle de f^{-1} en y est : $(\mathrm{d}f_{f^{-1}(y)})^{-1}$.*

Remarque. Notons $g = f^{-1}$. La différentielle de g est très facile à calculer si on différentie la relation : $f \circ g = \mathrm{Id}_W$, ce qui donne : $(\mathrm{d}f_{g(y)}) \circ (\mathrm{d}g_y) = (\mathrm{dId}_W)_y = \mathrm{Id}$.
Lorsque $n = 1$, on retrouve évidemment : $(f^{-1})'(y) = 1/f'(f^{-1}(y))$.

Une idée de preuve possible. C'est la méthode de Newton !
En dimension 1, on se donne y proche de $f(x^0)$ et on cherche à résoudre l'équation $f(x) = y$. Pour cela, on définit une suite par récurrence par :

$$x^{k+1} = x^k - \frac{f(x^k) - y}{f'(x^k)}.$$

On constate qu'en écrivant $g(x) = x - (f'(x))^{-1}(f(x) - y)$, g a un unique point fixe et, si f est $\mathcal{C}^2$, que g' s'annule en ce point fixe. C'est ce qui fait l'efficacité de la méthode de Newton : si on prend y assez proche de $f(x^0)$, la suite converge quadratiquement vers l'antécédent de y. Cela prouve l'existence de f^{-1}, reste à montrer que f^{-1} est dérivable, ce qu'on passe sous silence.
En dimension supérieure, on peut procéder de même, si ce n'est qu'il faut donner un sens à la suite. Puisque f est $\mathcal{C}1$, pour x assez proche de x^0, $\mathrm{d}f_x$ est inversible (les matrices inversibles forment un ouvert). Pour chercher un antécédent de y proche de $f(x^0)$, on définit une suite $(x^k)_{k\in\mathbb{N}}$ par :

$$x^{k+1} = x^k - (\mathrm{d}f_{x^k})^{-1}(f(x^k) - y)$$

et on montre qu'elle converge. Le reste est technique mais l'idée est claire.

6° Fonctions implicites

a) Observations préliminaires

Observons le cercle-unité C : globalement, ce n'est pas le graphe d'une fonction (de x...). Mais localement, à part au voisinage des points $(\pm 1, 0)$, c'en est un : pour tout point du cercle, il existe un voisinage de ce point autour duquel C est le graphe d'une fonction, celui de $\varphi : t \mapsto \varepsilon\sqrt{1 - t^2}$, où le signe ε est celui de x_2^0.

Plus formellement : pour $(x^0, y^0) \in C$, il existe V voisinage de (x^0, y^0), un réel $\alpha > 0$ et une fonction $\varphi : \left]x^0 - \alpha, x^0 + \alpha\right[\to \mathbb{R}$, tels que pour $(x, y) \in V$, $(x, y) \in C$ ssi $y = \varphi(x)$.
En général – c'est-à-dire si on remplace C par une courbe ayant une équation $f(x, y) = 0$ – on ne peut pas exprimer explicitement φ à partir de f. Le théorème des fonctions implicites permet de le faire... implicitement. Et ceci s'étend lorsque x et y ont eux-mêmes plusieurs composantes.

b) Énoncé

Théorème (Dimension 2)**.** *Soit U et V deux ouverts de $\mathbb{R}$. Soit $F : U \times V \to \mathbb{R}$ une fonction réelle de classe $\mathcal{C}^1$. Soit $(x^0, y^0) \in U \times V$ tel que $f(x^0, y^0) = 0$. On suppose que la dérivée partielle $\frac{\partial f}{\partial y}(x^0, y^0)$ n'est pas nulle. Alors, il existe un voisinage U' de x^0 et V' de y^0 et une fonction $\varphi : U' \to V'$ de classe $\mathcal{C}^1$ telle que*

$$\forall (x, y) \in U' \times V', \quad f(x, y) = 0 \iff y = \varphi(x).$$

De plus, on a :

$$\forall x \in U', \quad \varphi'(x) = -\frac{\frac{\partial f}{\partial x}(x, \varphi(x))}{\frac{\partial f}{\partial y}(x, \varphi(x))}.$$

Remarque. L'hypothèse du théorème des fonctions implicites est simple à retenir si on a l'image du cercle en tête : l'obstacle pour pouvoir exprimer y en fonction de x, c'est que la tangente soit verticale ; mais l'équation de la tangente, c'est $(x - x^0)\partial f/\partial x + (y - y^0)\partial f/\partial y = 0$: elle n'est pas verticale SSI $\partial f/\partial y \neq 0$.
Si f est une fonction affine de y, on peut tirer y de la relation $f(x, y) = 0$ exactement lorsque le coefficient de y, qui est précisément $\partial f/\partial y$, n'est pas nul.
La valeur de φ' se calcule aisément en dérivant la relation : $f(x, \varphi(x)) = 0$.

Théorème (Dimension $n+1$)**.** *Soit U un ouvert de $\mathbb{R}^n$ et V un ouvert de $\mathbb{R}$. Soit $F : U \times V \to \mathbb{R}$ une fonction réelle de classe $\mathcal{C}^1$. Soit $(x^0, y^0) \in U \times V$ tel que $f(x^0, y^0) = 0$. On suppose que la dérivée partielle $\frac{\partial f}{\partial y}(x^0, y^0)$ n'est pas nulle. Alors, il existe un voisinage U' de x^0 et V' de y^0 et une fonction $\varphi : U' \to V'$ de classe $\mathcal{C}^1$ telle que*

$$\forall (x, y) \in U' \times V', \quad f(x, y) = 0 \iff y = \varphi(x).$$

De plus, on a :

$$\forall x \in U', \ \forall i \in \{1, \ldots, n\}, \quad \frac{\partial \varphi}{\partial x_i}(x) = -\frac{\frac{\partial f}{\partial x_i}(x, \varphi(x))}{\frac{\partial f}{\partial y}(x, \varphi(x))}.$$

Exercice. Soit $f : \mathbb{R}^3 \to \mathbb{R}$ une fonction $\mathcal{C}^1$. On suppose qu'en un point $p = (x^0, y^0, z^0)$, on ait $f(p) = 0$ et les trois dérivées partielles de f sont non nulles. On peut donc, sur un voisinage de ce point, exprimer une variable en fonction des deux autres. On note par exemple $\left(\frac{\partial x}{\partial y}\right)_z$ la dérivée partielle par rapport à y de la fonction $x = \varphi(y, z)$.
Voici un exemple : $x = P$, $y = V$, $z = T$, n et R sont des constantes et $f(P, V, T) = PV - nRT$. On peut exprimer $P = nRT/V$, alors $\left(\frac{\partial x}{\partial y}\right)_z = -nRT/V^2$.
Démontrer (sur l'exemple et en général) que l'on a :

$$\left(\frac{\partial x}{\partial y}\right)_z \times \left(\frac{\partial y}{\partial z}\right)_x \times \left(\frac{\partial z}{\partial x}\right)_y = -1,$$

ce qui est un peu étonnant (on aurait attendu 1 en simplifiant...).

Théorème (Dimensions quelconques)**.** *Soit U un ouvert de $\mathbb{R}^n$ et V un ouvert de $\mathbb{R}^p$. Soit $F : U \times V \to \mathbb{R}^p$ une fonction de classe $\mathcal{C}^1$. Soit $(x^0, y^0) \in U \times V$ tel que $f(x^0, y^0) = 0$. On suppose que la matrice des dérivées partielles $\left(\frac{\partial f_i}{\partial y_j}(x^0, y^0)\right)$ est inversible. Alors, il existe un voisinage U' de x^0 et V' de y^0 et une fonction $\varphi : U' \to V'$ de classe $\mathcal{C}^1$ telle que*

$$\forall (x, y) \in U' \times V', \quad f(x, y) = 0 \iff y = \varphi(x).$$

De plus, on a :

$$\forall x \in U', \quad \mathrm{d}\varphi_x = -(\mathrm{d}f)_y^{-1} \circ (\mathrm{d}f)_x.$$

Remarque. L'hypothèse du théorème des fonctions implicites est simple à retenir si on a l'exemple où f est une application affine en tête. L'égalité $f(x, y) = 0$ est un système linéaire de p équations à p inconnues $y_1, \ldots, y_p$ et n paramètres $x_1, \ldots, x_n$. L'hypothèse est exactement que le déterminant de ce système n'est pas nul.

Exercice. Montrer que le théorème des fonctions implicites et le théorème d'inversion locale sont équivalents – comme ils sont tous les deux vrais, c'est évident : on veut dire ici qu'on peut les déduire relativement facilement l'un de l'autre...

Esquisse. Admettons le théorème d'inversion locale. Pour $(x, y) \in U \times V$, on pose

$$F(x, y) = (x, f(x, y))$$

L'hypothèse sur f fait que la différentielle de F est localement inversible (vérifier !). Le théorème d'inversion locale donne l'existence d'une fonction réciproque $G = F^{-1}$ définie sur un voisinage de $(x^0, 0)$. Alors, pour tout x au voisinage de x^0, $G(x, 0)$ est de la forme $(x, \varphi(x))$: l'application φ convient, essentiellement car $y = \varphi(x)$ SSI $(x, y) = G(x, 0)$ SSI $(x, 0) = F(x, y)$ SSI $f(x, y) = 0$.
Inversement, admettons le théorème des fonctions implicites. On définit F sur un voisinage de (x^0, y^0) dans $\mathbb{R}^{2n}$ par :

$$F \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(x_1, \ldots, x_n) - y_1 \\ \vdots \\ f_n(x_1, \ldots, x_n) - y_n \end{pmatrix}.$$

La matrice des dérivées partielles de F par rapport à $(x_1, \ldots, x_n)$ est précisément la matrice jacobienne de f, qui est supposée inversible. En appliquant à F le théorème des fonctions implicites, on peut donc exprimer x en fonction de y, c'est-à-dire trouver une fonction g définie au voisinage de y^0 et assez régulière telle que sur un voisinage de (x^0, y^0), on ait :

$$x = g(y) \iff F(x, y) = 0 \iff y = f(x).$$

c) Inversion globale

Redisons-le : si $n \geqslant 2$, l'inversibilité locale d'une fonction n'entraîne pas l'injectivité.

Proposition. *Soit U un ouvert de $\mathbb{R}^n$. Si une application $f : U \to \mathbb{R}^n$ de classe $\mathcal{C}^1$ a une différentielle inversible en tout point et si elle est* injective *(hypothèse globale), alors c'est un difféomorphisme de U sur $f(U)$.*

En effet, par injectivité, f est une bijection de U sur $f(U)$; par hypothèse, f est de classe $\mathcal{C}^1$; il n'y a plus qu'à vérifier que f^{-1} est de classe $\mathcal{C}^1$, ce qui est une propriété locale : pour $y \in f(U)$, le fait que f^{-1} soit $\mathcal{C}^1$ sur un voisinage de y résulte du théorème d'inversion locale appliqué à f et au point $f^{-1}(y)$.

7° Dérivées supérieures

a) Motivation

Si $n = 1$, une des motivations pour étudier la dérivée seconde est la position de la courbe par rapport à sa tangente. C'est particulièrement pertinent pour savoir si un point critique – point où la dérivée s'annule – est un extremum ou pas.
Supposons en effet f de classe $\mathcal{C}^2$ sur un voisinage de x et écrivons le DL à l'ordre 2 de f. Pour h assez petit, on a : $f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + o(h^2)$. Si $f''(x) \neq 0$, alors $f(x+h) - f(x) - hf'(x)$ est du signe de $f''(x)$ pour h assez peti. Si $f''(x) = 0$, on ne peut rien conclure du DL à l'ordre 2, observer le signe de $\Delta(h) = f(x+h) - f(x) - hf'(x)$ si $\Delta(h)$ vaut h^3 (la courbe traverse la tangente), $-h^4$ (la courbe reste sous la tangente), h^4 (la courbe reste au-dessus de la tangente).

b) Approche « abstraite »

Soit $U \subset \mathbb{R}^n$ et $f : U \to \mathbb{R}^m$. En un point x fixé, la différentielle en x est une application de $\mathbb{R}^n$ dans $\mathbb{R}^m$; on voit $\mathrm{d}f : x \mapsto \mathrm{d}f_x$ comme une application de U dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ – on peut regarder les matrices dans la base canonique, on associe à x la matrice $(a_{ij}(x))$ où $a_{ij} = \frac{\partial f_i}{\partial x_j}$. L'espace $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est de dimension finie $M = mn$, avec une base de matrices élémentaires E_{ij}, cela a un sens de voir si elle est différentiable. La différentielle de $\mathrm{d}f$ en x est, si elle existe, une application $\mathrm{d}^2 f$ de U dans $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$. On peut identifier $\mathrm{d}^2 f_x$ à une application bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}^m$: pour $h \in \mathbb{R}^n$, $\mathrm{d}^2 f_x(h)$ est une application qui à $h' \in \mathbb{R}^n$ associe $\mathrm{d}^2 f_x(h)(h') \in \mathbb{R}^m$; il revient au même de considérer que $\mathrm{d}^2 f_x$ associe un vecteur de $\mathbb{R}^m$ à tout couple de vecteurs $(h, h') \in \mathbb{R}^n \times \mathbb{R}^n$, et la condition de bilinéarité se vérifie facilement.
En termes d'éléments de matrices, de même que la différentielle de $f = (f_1, \ldots, f_m)$ était décrite par les dérivées partielles $\frac{\partial f_i}{\partial x_j}$, de même la différentielle seconde est décrite par les dérivées partielles $\frac{\partial a_{ij}}{\partial x_k}$. Coup de chance, la moitié ou presque des ces dérivées partielles ne servent à rien.

Lemme (Schwarz)**.** *Si toutes les dérivées partielles de f sont de classe $\mathcal{C}^1$, alors, pour tout $1 \leqslant i \leqslant m$ et tous $1 \leqslant j, k \leqslant n$:*

$$\frac{\partial}{\partial x_k}\frac{\partial f_i}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f_i}{\partial x_k}.$$

Ce lemme s'étend aux dérivées suivantes : si les dérivées partielles sont continues, l'ordre de dérivation n'importe pas.

c) Approche pragmatique

Pour simplifier, on suppose ici que $m = 1$. Soit donc $f : U \to \mathbb{R}$ pour U ouvert de $\mathbb{R}^n$. On dit que f est de classe $\mathcal{C}^p$ si toutes les dérivées partielles d'ordre p existent et sont continues. Par le lemme de Schwarz, on peut alors les calculer dans n'importe quel ordre.

Exemple. Supposons que $n = 2$ et que f soit un polynôme de degré $\leqslant 2$, c'est-à-dire :

$$f(x_1, x_2) = a_0 + a_{10}x_1 + a_{01}x_2 + a_{20}x_1^2 + a_{11}x_1x_2 + a_{02}x_2^2.$$

On étudie f au voisinage de $0 = (0, 0)$. Après s'être rappelé que $a_{10} = \frac{\partial f}{\partial x_1}(0)$ et $a_{01} = \frac{\partial f}{\partial x_2}(0)$, on calcule les dérivées secondes :

$$\frac{\partial^2 f}{\partial x_1^2}(0) = 2a_{20}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2}(0) = a_{11}, \quad \frac{\partial^2 f}{\partial x_2^2}(0) = 2a_{02}.$$

On a donc :

$$f(x_1,x_2) = f(0) + \frac{\partial f}{\partial x_1}(0)\cdot x_1 + \frac{\partial f}{\partial x_2}(0)\cdot x_2 + \frac{1}{2}\frac{\partial^2 f}{\partial x_1^2}(0)\cdot x_1^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}(0)\cdot x_1x_2 + \frac{1}{2}\frac{\partial^2 f}{\partial x_2^2}(0)\cdot x_2^2.$$

En introduisant la *hessienne* H de f en $0 = (0,0)$, on peut donc écrire :

$$f(x) = f(0) + \langle \nabla f_0, x\rangle + +\frac{1}{2}{}^t xHx, \quad \text{où } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ et } H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(0) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(0) & \frac{\partial^2 f}{\partial x_2^2}(0) \end{pmatrix}.$$

Pour un développement en un point quelconque x, on fait comme d'habitude un changement de variable pour se ramener en 0, c'est-à-dire que l'on étudie : $g(h) = f(x+h)$ pour $h \in \mathbb{R}^2$, fonction à laquelle on applique la formule précédente. On constate que $\frac{\partial g}{\partial x_i}(0) = \frac{\partial f}{\partial x_i}(x)$, etc., de sorte que le développement prend la forme suivante :

$$f(x+h) = f(h) + \langle \nabla f_x, h\rangle + +\frac{1}{2}{}^t hHh, \quad \text{où } h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \text{ et } H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{pmatrix}.$$

Exemple. Supposons que f soit un polynôme :

$$f(x_1,\dots,x_n) = \sum_{p=0}^{P} \sum_{i_1+\cdots+i_n=p} a_{i_1,\dots,i_n} x_1^{i_1}\cdots x_n^{i_n}.$$

Dans une dérivée partielle en $0 = (0,\dots,0)$, seul compte le terme dont l'exposant correspond à la dérivation :

$$\frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1}\cdots \partial x_n^{i_n}}(0) = i_1!\cdots i_n! \times a_{i_1,\dots,i_n}.$$

On a donc une formule de Taylor exacte pour les polynômes :

$$f(x_1,\dots,x_n) = \sum_{p=0}^{P} \sum_{i_1+\cdots+i_n=p} \frac{1}{i_1!\cdots i_n!} \frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1}\cdots \partial x_n^{i_n}}(0)\ x_1^{i_1}\cdots x_n^{i_n}.$$

Pour un développement en un point x, on étudie $g(h) = f(x+h)$ pour $h \in \mathbb{R}^n$, fonction à laquelle on applique la formule précédente. Le développement prend la forme suivante :

$$f(x+h) = \sum_{p=0}^{P} \sum_{i_1+\cdots+i_n=p} \frac{1}{i_1!\cdots i_n!} \frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1}\cdots \partial x_n^{i_n}}(x)\ h_1^{i_1}\cdots h_n^{i_n}.$$

Comme dans l'exemple précédent, les premiers termes de ce développement peuvent s'écrire :

$$f(x+h) = f(x) + \langle \nabla f_x, h\rangle + \frac{1}{2}\ {}^t hHh + \cdots,$$

où H est la *hessienne* de f en x, c'est-à-dire

$$H = \mathrm{Hess}(f)_x = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{1\leqslant i,j\leqslant n}.$$

Revenons au cas général : la définition précédente de hessienne a un sens et la formule de Taylor reste valable... à peu de chose près.

Théorème (Taylor-Young)**.** *Soit P un entier. Soit U un voisinage ouvert de $x \in \mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction admettant des dérivées partielles continues d'ordre P. Alors, pour $h \in \mathbb{R}^n$ assez petit, on a :*

$$f(x+h) = \sum_{p=0}^{P} \sum_{i_1+\cdots+i_n=p} \frac{1}{i_1!\cdots i_n!} \frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1}\cdots \partial x_n^{i_n}}(x)\ h_1^{i_1}\cdots h_n^{i_n} + o(||h||^P).$$

Si $P = 2$, on a en particulier :

$$f(x+h) = f(x) + \langle \nabla f_x, h\rangle + \frac{1}{2}\ {}^t hHh + o(||h||^2), \qquad \text{où } H = \mathrm{Hess}(f)_x = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{1\leqslant i,j\leqslant n}.$$

d) Application aux extrema

Soit U un ouvert et $f : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$.

Lemme. *Si f admet un extremum en $x \in U$, alors x est un point critique :* $\mathrm{d}f_x = 0$*.*

Démonstration. Soit v un vecteur non nul. La fonction $g :]-\varepsilon,\varepsilon[\to \mathbb{R}$, $t \mapsto g(x+tv)$ est bien définie pour ε assez petit et elle admet un extremum en 0, d'où : $\mathrm{d}f_x(v) = g'(0) = 0$. Comme c'est vrai pour tout v, il vient : $\mathrm{d}f_x = 0$.

Si f est $\mathcal{C}^2$, on peut discuter le signe de $f(x+h) - f(x)$ en fonction de la hessienne de f en x.

- Si la hessienne est définie positive, f admet un minimum local en x : en effet, soit $\lambda > 0$ la valeur minimale atteinte par la fonction $v \mapsto {}^t vHv$ sur la sphère unité (qui est compacte) : λ est la plus petite valeur propre de H. Alors, le DL de f s'écrit, en posant $v = h/||h||$:

  $$f(x+h) = f(x) + \frac{1}{2}\ {}^t hHh + o(||h||^2) = f(x) + \frac{1}{2}||h||^2({}^t vHv + o(1)) \geqslant f(x) + ||h||^2(\lambda + o(1)),$$

  et la fonction $\lambda + o(1)$ est strictement positive sur un voisinage de x.
- Si la hessienne est définie négative, f admet un maximum local en x.
- Si la hessienne possède deux valeurs propres non nulles de signes contraires, f n'admet pas d'extremum en x. En effet, soit v_+ (resp. v_-) un vecteur propre de H associé à la valeur propre $\lambda_+ > 0$ (resp. $\lambda_- < 0$). Considérons les fonctions $g_\pm :]-\varepsilon,\varepsilon[\to \mathbb{R}$, $t \mapsto f(x+tv_\pm)$. On a au voisinage de 0 : $g_\pm(t) - f(x) = t^2(\lambda_\pm/2 + o(1))$, qui est du signe de $\lambda_\pm$. L'existence de λ_+ (resp. λ_-) empêche f d'avoir un maximum (resp. un minimum) en x.
- Si la hessienne est dégénérée, on ne peut pas être sûr d'avoir un extremum, car cela dépend « du o », comme en témoignent les exemples suivants ($n = 2$, $x = 0$) :
  - si $f(x_1,x_2) = x_1^2 + x_1^3$, f admet un minimum en 0,
  - si $f(x_1,x_2) = x_1^2 + x_2^4$, f admet un minimum en 0,
  - si $f(x_1,x_2) = x_1^2 + x_2^3$, f n'admet pas d'extremum en 0,
  - si $f(x_1,x_2) = x_1^2 - x_2^4$, f n'admet pas d'extremum en 0, etc.

Remarque. Ce qui précède permet, plus généralement, de déterminer la position du graphe par rapport au plan tangent, *i.e.* le signe de $f(x+h) - f(x) - \mathrm{d}f_x(h)$.