Quelques rappels sur le calcul différentiel

On munit dans tout ce qui suit les espaces de leurs normes euclidiennes canoniques. On note $\mathscr{L}(\mathbf{R}^n;\mathbf{R}^p)$ l'ensemble des applications linéaires (continues) de $\mathbf{R}^n$ dans $\mathbf{R}^p$. Enfin on identifie les vecteurs de $\mathbf{R}^n$, $n \in \mathbf{N}^*$, avec des vecteurs colonnes. Par ailleurs, on rappelle qu'une application qui s'écrit comme somme d'une application constante et d'une application linéaire est dite affine.

1 Différentier

1.1 Préliminaires

De manière informelle : pour une fonction raisonnable, on s'attend à ce qu'autour d'un point donné, une petite variation au départ corresponde essentiellement à une petite variation proportionnelle à l'arrivée. La bonne définition pour « proportionnelle » est la linéarité.
Évidemment, même localement, une application, même raisonnable, n'est affine qu'approximativement. Pour rendre cela précis, on va rappeler ce que veulent dire localement et approximativement.

Définition 1 *On dit qu'une propriété est vraie* au voisinage d'un point $A \in \mathbf{R}^n$ *s'il existe une boule (non triviale) centrée en A telle que la propriété soit vraie en tout point de la boule.*

Remarquons que pour $n = 1$ une boule est en fait un intervalle.
Pour définir ce qui est néligigeable et ce qui ne l'est pas, il faut savoir comparer des fonctions.

Définition 2 Notations de Landau.
Soit $f : \Omega \to \mathbf{R}^p$ et $g : \Omega \to \mathbf{R}$ (où $\Omega \subset \mathbf{R}^n$).

1. *On dit que f est* négligeable *devant g au voisinage de $A \in \Omega$ s'il existe une fonction $\varepsilon : \Omega \to \mathbf{R}^p$* tendant vers 0 *en A et telle qu'au voisnage de A l'on ait $\forall X \in \Omega$, $f(X) = g(X)\varepsilon(X)$. On note alors*
$$f(X) \stackrel{A}{=} o(g(X)).$$
2. *On dit que f est* dominée *par g au voisinage de $A \in \Omega$ s'il existe une fonction $M : \Omega \to \mathbf{R}^p$* bornée *telle qu'au voisinage de A l'on ait $\forall X \in \Omega$, $f(X) = g(X)M(X)$. On note alors*
$$f(X) \stackrel{A}{=} \mathcal{O}(g(X)).$$

Remarque 3 *Une fois que l'on a négligé ou borné un terme, on ne peut plus revenir en arrière.*

Exemple 4 *$f(X) \stackrel{A}{=} o(1)$ signifie que f tend vers 0 en A.*
En particulier f est continue en A si $f(X) \stackrel{A}{=} f(A) + o(1)$.

Exemples 5 *Avec $n = 1$: $x^2 \stackrel{0}{=} o(x)$, $\sin(x) \stackrel{0}{=} \mathcal{O}(1)$.*

1.2 Différentielle

Définition 6 *Soit $f : \Omega \to \mathbf{R}^p$ (où $\Omega \subset \mathbf{R}^n$) et $A \in \Omega$ tel que Ω contienne un voisinage de A.*

1. *On dit que f est* différentiable *en A s'il existe $\ell : \mathbf{R}^n \to \mathbf{R}^p$ linéaire (continue) telle que*

$$f(X) \overset{A}{=} f(A) + \ell(X - A) + o(\|X - A\|).$$

On note alors $\ell = df(A)$. La matrice associée $Jac(f)(A)$ est appelée matrice Jacobienne *ou simplement Jacobienne.*

2. *Quand f est différentiable en tout point, on dit simplement que f est différentiable et l'on note $df : \Omega \to \mathcal{L}(\mathbf{R}^n; \mathbf{R}^p)$ sa différentielle.*

Remarque 7 *Il est souvent plus commode de se ramener en $(0, \cdots, 0)$:*

$$f(A + H) \overset{(0,\cdots,0)}{=} f(A) + df(A)(H) + o(\|H\|).$$

Remarque 8 *Quand $n = 1$, on retrouve la notion de dérivabilité avec $df(a) : \mathbf{R} \to \mathbf{R}^p$, $h \mapsto h\, f'(a)$.*

Remarque 9 *Un autre cas particulier intéressant est celui où $p = 1$, car toute application linéaire de $\mathbf{R}^n$ vers $\mathbf{R}$ (appelée* forme linéaire*) s'écrit comme un produit scalaire avec un certain vecteur. Le vecteur correspondant à la différentielle est appelé* gradient *et noté $\nabla f(A)$, il est relié à la différentielle par*

$$df(A)(H) \;=\; \langle \nabla f(A), H \rangle.$$

Observation 10 *L'interprétation en terme de produit scalaire donne des informations géométriques intuitives : si l'on veut augmenter la valeur de f il faut se déplacer dans la direction du gradient, si l'on veut diminuer la valeur de f il faut se déplacer dans la direction opposée, si l'on veut ne pas changer f il faut se déplacer perpendiculairement. Ainsi les lignes de niveaux (ensemble de points prenant la même valeur) sont en tout point orthogonales au gradient. De plus, si f atteint sur un voisinage de A son minimum (ou son maximum) en A (on dit que f atteint un minimum (ou maximum) local), alors $\nabla f(A) = (0, \cdots, 0)$, on dit que A est un* point critique *de f.*

Exemple 11 $f : \mathbf{R}^2 \to \mathbf{R}$, $(x, y) \longmapsto x\, y$.

$$\begin{aligned} f(A + H) &= (x_A + x_H)\,(y_A + y_H) \\ &= x_A\, y_A + x_H\, y_A + x_A\, y_H + x_H\, y_H \\ &= x_A\, y_A + x_H\, y_A + x_A\, y_H + o(\|(x_H, y_H)\|). \end{aligned}$$

Donc l'application f est différentiable en $A = (x_A, y_A)$ et sa différentielle en A, $df(A) : \mathbf{R}^2 \to \mathbf{R}$, est donnée par, pour $H = (x_H, y_H)$, $df(A)(H) = x_H\, y_A + x_A\, y_H$. Le gradient en A est donc $\nabla f(A) = (y_A, x_A)$.

Exemple 12 *L'exemple précédent se généralise immédiatement à toutes les applications bilinéaires et plus généralement aux applications multilinéaires.*
Si $B : (\mathbf{R}^n)^k \to \mathbf{R}^p$ est une application k-linéaire (continue) et que l'on définit $f : \mathbf{R}^n \to \mathbf{R}^p$, $X \mapsto B(X, \cdots, X)$, alors f est différentiable et sa différentielle en un point A est donnée par, pour tout H,

$$df(A)(H) \;=\; \sum_{j=1}^{k} B(\underbrace{A, \cdots, A}_{j-1}, H, \underbrace{A, \cdots, A}_{k-j}).$$

Dans le cas particulier où $p=1$, $k=2$ et B est le produit scalaire, alors f est le carré de la norme et l'on obtient $df(A)(H) = \langle A, H\rangle + \langle H, A\rangle = 2\langle A, H\rangle$ donc $\nabla f(A) = 2A$.

Exemple 13 *Soit $d \in \mathbf{N}^*$. Posons $\Omega = Gl_d(\mathbf{R})$ l'ensemble des matrices inversibles ($n = p = d^2$) et définissons*

$$f : Gl_d(\mathbf{R}) \to Gl_d(\mathbf{R}), \quad A \mapsto A^{-1}.$$

Pour tout $A \in Gl_d(\mathbf{C})$ et tout $H \in \mathcal{M}_d(\mathbf{C})$ tel que $\|H\| < 1$, on a

$$\begin{aligned} f(A+H) &= (A+H)^{-1} = A^{-1}(I_d + H\,A^{-1})^{-1} \\ &= A^{-1}\sum_{k=0}^{+\infty}(-HA^{-1})^k \\ &\overset{O_d}{=} A^{-1} - A^{-1}\,H\,A^{-1} + o(\|H\|)\,. \end{aligned}$$

D'où $df(A)(H) = -A^{-1}\,H\,A^{-1}$. Quand $d=1$, on retrouve $f'(a) = -a^{-2}$.

Exemple 14 *$A = (x_A, y_A)$ et $f : \mathbf{R}^2 \to \mathbf{R}^3$, $(x,y) \mapsto (y^2, \cos(x+y), x\,e^y)$. f est différentiable en A et $df(A) : \mathbf{R}^2 \to \mathbf{R}^3$ est donné par, pour $H = (x_H, y_H)$,*

$$\begin{aligned} df(A)(H) &= \begin{bmatrix} 0 & 2y_A \\ -\sin(x_A+y_A) & -\sin(x_A+y_A) \\ e^{y_A} & x_A\,e^{y_A} \end{bmatrix} \begin{pmatrix} x_H \\ y_H \end{pmatrix} \\ &= (2y_A\,y_H, -\sin(x_A+y_A))\,(x_H+y_H), e^{y_A}\,x_H \;+\; x_A\,e^{y_A}\,y_H)\,. \end{aligned}$$

1.3 Dérivées partielles

Comme le laisse penser ce dernier exemple, il peut être plus commode de calculer en faisant varier coordonnée par coordonnée, en dérivant direction par direction.

Définition 15 *Soit $f : \Omega \to \mathbf{R}^p$ (où $\Omega \subset \mathbf{R}^n$) et $A \in \Omega$ tel que Ω contienne un voisinage de A.*

1. *On dit que f est dérivable en $A \in \Omega$ dans la direction H si la fonction $I \to \mathbf{R}^p$, $t \mapsto f(A + t\,H)$ (où I est un intervalle ouvert centré en 0) est dérivable en 0. Cette* dérivée directionnelle *est alors notée $\partial_H f(A) = \dfrac{\partial f}{\partial H}(A)$.*
2. *Quand H est l'un des vecteurs de la base (canonique) e_i correspondant à une coordonnée X_i, on parle plutôt de* dérivée partielle *et l'on note $\partial_i f(A) = \partial_{X_i} f(A) = \dfrac{\partial f}{\partial X_i}(A)$.*

Théorème 16
1. *Si f est différentiable en A, alors f y possède des dérivées dans toutes les directions, $\partial_H f(A) = df(A)(H)$ et*

$$df(A)(H) = \frac{\partial f}{\partial X_1}(A)\;H_1 + \cdots + \frac{\partial f}{\partial X_n}(A)\;H_n.$$

En particulier, alors $\frac{\partial f}{\partial X_i}(A) = df(A)(e_i)$.

2. *Réciproquement, si f possède dans un voisinage dans A des dérivées partielles (par rapport à toutes les coordonnées) continues, alors f possède une différentielle en A.*

Démonstration. Donnons une idée de la démonstration de la réciproque quand $n = 2$ (pour simplifier les notations) :

$$\begin{aligned}
f(A+H) - f(A) &= f(A + H_1e_1 + H_2e_2) - f(A + H_1e_1) + f(A + H_1e_1) - f(A) \\
&\overset{(0,0)}{=} \frac{\partial f}{\partial X_2}(A + H_1e_1)\, H_2 + o(H_2) + \frac{\partial f}{\partial X_1}(A)\, H_1 + o(H_1) \\
&\overset{(0,0)}{=} \frac{\partial f}{\partial X_1}(A)\, H_1 + \frac{\partial f}{\partial X_2}(A)\, H_1 + o(\|H\|).
\end{aligned}$$

■

Observation 17 *En remarquant que les différentielles des coordonnées sont données par* $dX_i(A)(H) = H_i$*, on a donc*

$$df(A) = \sum_i \frac{\partial f}{\partial X_i}(A)\, dX_i(A)$$

ou encore

$$df(A)(H) = \sum_i \frac{\partial f}{\partial X_i}(A)\, dX_i(A)(H) = \sum_i \frac{\partial f}{\partial X_i}(A)\, H_i.$$

Si f *est différentiable on a donc* $df = \sum_i \frac{\partial f}{\partial X_i} dX_i$*. Avec* $n = 1$*, on a en particulier* $df = f'\,dx$*.*

Observation 18 *Cela implique qu'en écrivant* f *en coordonnées* $f = (f_1, \cdots, f_p)$*,*

$$[Jac(f)(A)]_{i,j} \;=\; \frac{\partial f_i}{\partial X_j}(A)\,.$$

Quand $p = 1$*, on trouve aussi*

$$\nabla f(A) \;=\; \left(\frac{\partial f}{\partial X_1}(A), \cdots, \frac{\partial f}{\partial X_n}(A)\right)\,.$$

Définition 19 *Soit* $\Omega \subset \mathbf{R}^n$ *un ouvert. On note* $\mathcal{C}^1(\Omega, \mathbf{R}^p)$*, ou simplement* $\mathcal{C}^1(\Omega)$ *si* $p = 1$*, l'ensemble des fonctions de* Ω *dans* $\mathbf{R}^p$ *différentiables à différentielle continue. De telles fonctions sont dites* de classe $\mathcal{C}^1$ *ou continûment différentiable.*

1.4 Opérations

Proposition 20 *Si*
- $\Omega \subset \mathbf{R}^n$ *est un voisinage de* A*,* $f : \Omega \to \mathbf{R}^p$ *est différentiable en* A*,*
- $\Omega' \subset \mathbf{R}^p$ *est un voisinage de* $f(A)$ *et* $g : \Omega' \to \mathbf{R}^q$ *est différentiable en* $f(A)$*,*

alors $g \circ f$ *est différentiable en* A *et* $d(g \circ f)(A) \;=\; dg(f(A)) \circ df(A)$*, c'est-à-dire*

$$d(g \circ f)(A)(H) \;=\; dg(f(A))(df(A)(H)).$$

Démonstration. Cela se démontre avec

$$\begin{aligned}
(g \circ f)(A+H) &\overset{(0,\cdots,0)}{=} g(f(A) + \mathrm{d}f(A)(H) + o(\|H\|)) \\
&\overset{(0,\cdots,0)}{=} g(f(A)) + \mathrm{d}g(f(A))(\mathrm{d}f(A)(H) + o(\|H\|)) + o(\|\mathrm{d}f(A)(H) + o(H)\|) \\
&\overset{(0,\cdots,0)}{=} g(f(A)) + \mathrm{d}g(f(A))(\mathrm{d}f(A)(H)) + o(\|H\|).
\end{aligned}$$

■

Observation 21 *En coordonnées cela se récrit*

$$\partial_i(g \circ f)(A) \;=\; \sum_{j=1}^{n} \partial_j g(f(A))\,\partial_i f_j(A)\,.$$

Matriciellement cela se récrit

$$Jac(g \circ f)(A) \;=\; Jac(g)(f(A))\,Jac(f)(A)\,.$$

Observation 22 *Cela permet de montrer beaucoup de relations. On pourra en particulier considérer le cas où g est une application multilinéaire.*

1.5 Ordre supérieur et développement de Taylor

Lorsque f est différentiable (en tout point de Ω), la différentielle $\mathrm{d}f$ étant elle-même une fonction, on peut vouloir la différentier. Quand c'est possible, on dit que f est *deux fois différentiable* en A et l'on introduit la notation $\mathrm{d}^2 f(A)$ (définissant une application bi-linéaire) pour la différentielle seconde au point A

$$\mathrm{d}f(A+H') \overset{(0,\cdots,0)}{=} \mathrm{d}f(A) + \mathrm{d}^2 f(A)(\cdot, H') + o(\|H'\|)$$

c'est-à-dire, pour tout H,

$$\mathrm{d}f(A+H')(H) \overset{(0,\cdots,0)}{=} \mathrm{d}f(A)(H) + \mathrm{d}^2 f(A)(H, H') + o(\|H'\|).$$

Quand $n = 1$, on trouve $\mathrm{d}f^2(a)(h, h') = h\,h'\,f''(a)$.
En sachant comment varie la différentielle, on obtient une information plus précise sur f.

Proposition 23 Développement de Taylor à l'ordre 2. *Si $f : \Omega \to \mathbf{R}^p$ est deux fois différentiable en A, alors*

$$f(A+H) \overset{(0,\cdots,0)}{=} f(A) + df(A)(H) + \frac{1}{2} d^2 f(A)(H,H) + o(\|H\|^2)\,.$$

Remarque 24 *Contrairement au cas des dérivées premières, il n'y a pas de réciproque.*

Corollaire 25 *Si $f : \Omega \to \mathbf{R}$ ($p = 1$) est deux fois différentiable en A, que A est un point critique (c'est-à-dire que $df(A)$ est nulle) et que $d^2 f(A)$ est définie positive, alors f atteint un minimum local strict en A.*

Observation 26 *La formule est exacte quand f est quadratique. En effet si $B : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^p$ est bilinéaire et $f : \mathbf{R}^n \to \mathbf{R}^p$, $X \to B(X, X)$, alors d'une part*

$$f(A+H) = B(A+H, A+H) = B(A,A) + B(A,H) + B(H,A) + B(H,H)$$

et d'autre part $df(A)(H) = B(A,H) + B(H,A)$, $d^2 f(A)(H,H') = B(H',H) + B(H,H')$.

On peut évidemment passer par les dérivées partielles.
Si f possède (dans un voisinage de A) une dérivée dans la direction H' et que cette dérivée partielle est dérivable dans la direction H en A, on note

$$\frac{\partial^2 f}{\partial H \partial H'}(A) \; = \; \frac{\partial}{\partial H}\left(\frac{\partial f}{\partial H'}\right)(A).$$

Si $i = j$, on note aussi

$$\frac{\partial^2 f}{\partial H \partial H}(A) \; = \; \frac{\partial^2 f}{\partial H^2}(A)\,.$$

Comme pour les dérivées d'ordre 1, si $H = e_i$ et $H' = e_j$, on parle de dérivée partielle et l'on note

$$\frac{\partial^2 f}{\partial X_i \partial X_j}(A)\,, \quad \text{et} \quad \frac{\partial^2 f}{\partial X_i^2}(A) \quad \text{si } i = j\,.$$

Corollaire 27 *1. Si f est deux fois différentiable en A, alors f y possède des dérivées partielles d'ordre 2 dans toutes les directions $\frac{\partial^2 f}{\partial H \partial H'}(A) = d^2 f(A)(H, H')$ et*

$$d^2 f(A)(H, H') \; = \; \sum_{i,j} \frac{\partial^2 f}{\partial X_j \partial X_i}(A)\; H_i\, H'_j\,.$$

En particulier,

$$\frac{\partial^2 f}{\partial X_j \partial X_i}(A) \; = \; d^2 f(A)(e_i, e_j)\,.$$

2. Réciproquement, si f possède dans un voisinage de A des dérivées partielles d'ordre 2 (par rapport à toutes les coordonnées) continues, alors f possède une différentielle d'ordre 2 en A,

Théorème 28 Théorème de Schwarz. *Si f possède dans un voisinage dans A des dérivées partielles d'ordre 2 (par rapport à toutes les coordonnées) continues, alors $d^2 f(A)$ est symétrique, c'est-à-dire*

$$d^2 f(A)(H, H') \; = \; d^2 f(A)(H', H).$$

En particulier,

$$\frac{\partial^2 f}{\partial X_i \partial X_j}(A) \; = \; \frac{\partial^2 f}{\partial X_j \partial X_i}(A).$$

La relation de Schwarz exprime que pour une fonction raisonnable un petit accroissement H suivi d'un petit accroissement H' a (approximativement) le même effet qu'un petit accroissement de H' suivi d'un petit accroissement H.

Remarque 29 *Quand $p = 1$, on introduit la* matrice Hessienne *$Hess(f)(A)$ la matrice telle que*

$$d^2 f(A)(H, H') \; = \; \langle H, Hess(f)(A)H' \rangle.$$

En particulier, $[Hess(f)(A)]_{i,j} = \dfrac{\partial^2 f}{\partial X_j \partial X_i}(A)$.

Définition 30 *Soit $\Omega \subset \mathbf{R}^n$ un ouvert. On note $\mathcal{C}^2(\Omega, \mathbf{R}^p)$, ou simplement $\mathcal{C}^2(\Omega)$ si $p = 1$, l'ensemble des fonctions de Ω dans $\mathbf{R}^p$ deux fois différentiables à différentielle seconde continue. De telles fonctions sont dites* de classe $\mathcal{C}^2$ *ou deux fois continûment différentiable.*

Évidemment, on peut obtenir des équivalents avec plus de différentiations...

2 Du local au global

2.1 Développements de Taylor

On a énoncé ci-dessus les formules de Taylor-Young. Pour obtenir une version plus précise, il faut se ramener en dimension 1 en choisissant un chemin $\gamma : [0,1] \to \Omega$ tel que $\gamma(0) = A$ et $\gamma(1) = B$ puis en développant $f(B) - f(A) = f \circ \gamma(1) - f \circ \gamma(0)$.
Évidemment le chemin le plus simple est le segment $[A, B]$ paramétré par

$$\gamma : [0,1] \to \Omega, \quad t \mapsto A + t\,(B - A) = (1-t)\,A + t\,B\,.$$

Avec ce choix, si f est k fois différentiable, on a, pour tout $t \in [0,1]$,

$$(f \circ \gamma)^{(k)}(t) \;=\; \mathrm{d}^k f(\gamma(t))\,(B - A, \cdots, B - A)\,.$$

D'où ce qui suit.

Théorème 31 *1.* **Développement de Taylor avec reste intégral.**
Soit $\Omega \subset \mathbf{R}^n$ un ouvert et $k \in \mathbf{N}$.
Si $f : \Omega \to \mathbf{R}^p$ est $(k+1)$ fois continûment différentiable et $[A, B] \subset \Omega$, alors

$$\begin{aligned} f(B) \;&=\; \sum_{j=0}^{k} \frac{1}{j\,!} d^j f(A)\,(B - A, \cdots, B - A) \\ &+\; \frac{1}{k\,!} \int_0^1 (1-t)^k\, d^{k+1} f(A + t(B - A))\,(B - A, \cdots, B - A)\, dt\,. \end{aligned}$$

2. **Formule de Taylor-Lagrange.**
Soit $\Omega \subset \mathbf{R}^n$ un ouvert et $k \in \mathbf{N}$.
Si $f : \Omega \to \mathbf{R}$ ($p = 1$) est $(k+1)$ fois différentiable et $[A, B] \subset \Omega$, alors il existe $\xi \in [A, B]$ tel que

$$\begin{aligned} f(B) \;&=\; \sum_{j=0}^{k} \frac{1}{j\,!} d^j f(A)\,(B - A, \cdots, B - A) \\ &+\; \frac{1}{(k+1)\,!} d^{k+1} f(\xi)\,(B - A, \cdots, B - A)\,. \end{aligned}$$

Remarque 32 *Dans les deux cas, on déduit*

$$\begin{aligned} \Big\| f(B) \;&-\; \sum_{j=0}^{k} \frac{1}{j\,!} d^j f(A)\,(B - A, \cdots, B - A) \Big\| \\ &\leq\; \frac{1}{(k+1)\,!} \left(\sup_{[A,B]} \| d^{k+1} f \| \right)\; \|B - A\|^{k+1}\,. \end{aligned}$$

Mais les hypothèses pour la formule de Taylor-Lagrange n'implique pas que la borne supérieure ci-dessus est finie.

Remarque 33 *Un ensemble Ω tel que $[A, B] \subset \Omega$ pour toute paire $(A, B) \in \Omega^2$ est dit* convexe. *Les convexes de $\mathbf{R}$ ($n = 1$) sont précisément les intervalles.*

2.2 Convexité

Quand $p = 1$, on peut chercher des minorations ou des majorations plutôt que des bornes.

Définition 34 *Soit $\Omega \subset \mathbf{R}^n$ un convexe et $f : \Omega \mapsto \mathbf{R}$.*

1. *On dit que f est* convexe *si, pour toute paire $(A, B) \in \Omega^2$, pour tout $t \in [0, 1]$,*
$$f((1-t)\,A + t\,B)) \;\leq\; (1-t)\,f(A) \;+\; t\,f(B)\,.$$
On dit que f est concave *si, pour toute paire $(A, B) \in \Omega^2$, pour tout $t \in [0, 1]$,*
$$f((1-t)\,A + t\,B)) \;\geq\; (1-t)\,f(A) \;+\; t\,f(B)\,.$$
2. *On dit que f est* strictement convexe *si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$, pour tout $t \in]0, 1[$,*
$$f((1-t)\,A + t\,B)) \;<\; (1-t)\,f(A) \;+\; t\,f(B)\,.$$
On dit que f est strictement concave *si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$, pour tout $t \in]0, 1[$,*
$$f((1-t)\,A + t\,B)) \;>\; (1-t)\,f(A) \;+\; t\,f(B)\,.$$

Remarque 35 *f est concave (respectivement strictement concave) si et seulement si $-f$ est convexe (respectivement strictement convexe). On ne discutera donc que la convexité.*

Proposition 36 *Soit $\Omega \subset \mathbf{R}^n$ un ouvert convexe et $f : \Omega \mapsto \mathbf{R}$.*

1. *Si f est différentiable, alors f est convexe si et seulement si, pour toute paire $(A, B) \in \Omega^2$,*
$$f(B) \;\geq\; f(A) \;+\; df(A)\,(B-A)\,.$$
2. *Si f est différentiable, alors f est strictement convexe si et seulement si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$,*
$$f(B) \;>\; f(A) \;+\; df(A)\,(B-A)\,.$$
3. *Si f est deux fois différentiable, alors f est convexe si et seulement si, en tout point $A \in \Omega$, $d^2f(A)$ est symétrique positive.*
4. *Si f est deux fois différentiable, alors si, en tout point $A \in \Omega$, $d^2f(A)$ est symétrique définie positive, f est strictement convexe.*

Démonstration. On peut le déduire directement du cas $n = 1$. Pour une démonstration *ab initio* : les points 1. et 2. découlent de simples manipulations algébriques, les points 3. et 4. sont des conséquences directes de la formule de Taylor-Lagrange à l'ordre 2. ∎

3 Inversibilité

On discute maintenant de comment convertir l'inversibilité du linéarisé en inversibilité (locale) non linéaire. Comme une majorité des résultats d'existence et d'unicité, la réponse est une conséquence du théorème de point fixe de Banach.

3.1 Point fixe

Définition 37 *Soit (X, d) un espace métrique.*

1. *Une suite $(u_n)_{n\in\mathbf{N}} \in X^{\mathbf{N}}$ est dite* de Cauchy *si*
$$\sup_{\min(\{p,q\})\geq n} d(u_p, u_q) \overset{n\to\infty}{\to} 0\,.$$
2. *(X, d) est dit* complet *si toutes ses suites de Cauchy convergent.*

Remarque 38 *Rappelons que tout suite convergente est de Cauchy.*

Remarque 39 *Un espace normé complet est appelé* espace de Banach. *L'exemple le plus utile d'espace métrique complet est fourni par les fermés des espaces de Banach.*

Définition 40 *Soit (X, d) et (Y, d') deux espaces métriques.*

1. *Pour $K \in \mathbf{R}_+$, une application $f : X \to Y$ est dite* K-Lipschitzienne *si, pour toute paire $(x, x') \in X^2$, on a*
$$d'(f(x), f(y)) \ \leq\ K\, d(x, y)\,.$$
2. *Une application $f : X \to Y$ est dite* Lipschitzienne *s'il existe $K \in \mathbf{R}_+$ tel que f soit K-Lipschitzienne.*
3. *Une application $f : X \to Y$ est dite* strictement contractante *s'il existe $K \in [0, 1[$ tel que f soit K-Lipschitzienne.*

Remarque 41 *Comme on l'a déjà vu en dimension finie pour les applications affines, être Lipschitzien est préservé par un changement de distance équivalente, mais pas être strictement contractant.*

Théorème 42 Théorème du point fixe de Banach *ou* **Théorème du point fixe de Picard** *ou* **Théorème de l'application contractante.**
Soit (X, d) un espace métrique complet non vide, $K \in [0, 1[$ et $\Phi : X \to X$ K-Lipschitzienne.

1. *Φ possède un unique point fixe.*
2. *Pour tout $x_0 \in X$, la suite $(x^{(k)})_{k\in\mathbf{N}}$ définie par $x^{(0)} = x_0$ et $(\forall k \in \mathbf{N},\ x^{(k+1)} = \Phi(x^{(k)}))$ converge vers l'unique point fixe de Φ et, de plus, en notant x_* ce point fixe, on a*
$$\forall k \in \mathbf{N}^*,\ d(x^{(k)}, x_*) \ \leq\ \frac{K^k}{1-K}\, d(\Phi(x_0), x_0)$$
et
$$\forall k \in \mathbf{N},\ d(x^{(k)}, x_*) \ \leq\ K^k\, d(x_0, x_*)\,.$$
La suite $(x^{(k)})_{k\in\mathbf{N}}$ est appelée suite des itérées de Picard.

Remarque 43 *C'est un théorème constructif. On peut approcher x_* par les itérées de Picard.*

Remarque 44 *La première estimation a l'avantage de ne pas dépendre de x_*. On peut donc l'utiliser pour arrêter un algorithme d'évaluation de x_*.*

Démonstration. **Au tableau.** ∎

3.2 Inversion locale

Pour espérer une forme d'inversibilité, on se restreint au cas $n = p$. Le cas $n > p$ peut être intégré à une version à paramètres du théorème qui suit et qui s'appelle alors *théorème des fonctions implicites.*

Théorème 45 Théorème d'inversion locale.
Soit $\Omega \subset \mathbf{R}^n$ un ouvert et $f : \Omega \mapsto \mathbf{R}^n$ de classe $\mathcal{C}^1$.
Soit $A \in \Omega$ tel que $df(A)$ soit inversible. Alors il existe un voisinage ouvert $\Omega' \subset \Omega$ de A et un voisinage ouvert U de $f(A)$ tel que

$$\Omega' \to U, \;\; M \mapsto f(M)$$

soit un difféomorphisme de classe $\mathcal{C}^1$, c'est-à-dire une bijection $\mathcal{C}^1$ de réciproque $\mathcal{C}^1$.
Si de plus f est de classe $\mathcal{C}^k$ pour un certain $k \in \mathbf{N}^*$, alors quitte à changer Ω' et U, on peut assurer que la réciproque l'est également.

Remarque 46 *Quand $n = 1$, la condition d'inversibilité de $df(a) : \mathbf{R} \to \mathbf{R}$, $m \to m\, f'(a)$ se récrit simplement $f'(a) \neq 0$.*

Remarque 47 *Les règles de différentiation des composées assure que la réciproque locale g ainsi construite vérifie en tout $B \in U$*

$$dg(B) \;=\; [df(g(B))]^{-1}\,.$$

Démonstration. Quelques éléments **au tableau.** ∎

Remarque 48 *Le théorème assure que localement on peut passer de l'inversibilité de l'application affine*

$$\mathbf{R}^n \to \mathbf{R}^n, \quad M \mapsto f(A) + df(A)(M - A)$$

à celle de f. On peut d'ailleurs montrer plus directement encore le théorème en appliquant le théorème de point fixe dans une boule fermée de $\mathcal{C}^1(U, \mathbf{R}^n)$ centrée en l'inverse de cette application affine,

$$\mathbf{R}^n \to \mathbf{R}^n, \quad B \mapsto A + [df(A)]^{-1}(B - f(A))\,.$$

Remarque 49 *Posons $B = f(A)$. Quitte à changer Ω' et U, le théorème assure en particulier que la seule solution $M \in \Omega'$ de $f(M) = B$ est $M = A$ et que, pour tout $B' \in U$, l'unique solution $A' \in \Omega'$ de $f(A') = B'$ vérifie*

$$\|A' - A\| \;\leq\; \left(\sup_{\xi \in \Omega'} |||[df(\xi)]^{-1}|||\right) \; \|B' - B\|\,.$$

Quand $n = 1$, l'estimation se récrit

$$|A' - A| \;\leq\; \frac{1}{\inf_{\Omega'} |f'|} \; |B' - B|\,.$$