Théorème d'inversion locale et théorème des fonctions implicites

Math 3.1. Calcul différentiel

Il est vivement conseillé d'apprendre soigneusement ce qui a déjà été fait en cours, et de faire les exercices qui n'ont pas encore été corrigés, avant d'essayer de lire ce qui suit.

1 Un exemple

Considérons la fonction $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ définie par

$$f(x, y) = x^2 + y^2 - 1\,.$$

Le graphe de cette fonction est alors une surface de $\mathbb{R}^3$ (cf. figure). Résoudre l'équation $f(x, y) = 0$ revient donc à chercher l'intersection de ce graphe avec avec le plan d'équation $z = 0$. L'ensemble des solutions est alors le cercle $\mathcal{C}$ d'équation

$$x^2 + y^2 = 1\,,$$

c'est-à-dire le cercle de centre O et de rayon 1. La question est de savoir si au voisinage d'un point (a, b) du cercle $\mathcal{C}$, celui-ci est le graphe d'une application φ, définie sur un ouvert de $\mathbb{R}$ et à valeurs $\mathbb{R}$. On voit que cela n'est pas possible pour les 2 points de coordonnées $(1, 0)$ et $(-1, 0)$. Cela est dû au fait que $\partial f/\partial y(\pm 1, 0) = 0$. Par contre, si (a, b) est un point du cercle $\mathcal{C}$ différent de ces deux derniers, alors on peut trouver un voisinage du point (a, b) qui n'intersecte pas l'axe des abscisses. Pour tout (x, y) appartenant à un tel voisinage,

$$\frac{\partial f}{\partial y}(x, y) \neq 0\,.$$

De plus, sur un tel voisinage l'équation du cercle peut se mettre sous la forme

$$y = \varphi(x) = \operatorname{signe}(b)\sqrt{1 - x^2}\,.$$

On voit alors que φ est dérivable sur ce voisinage et que la dérivée vérifie

$$\varphi'(x) = -\operatorname{signe}(b)\frac{x}{\sqrt{1 - x^2}} = -\frac{\partial f/\partial x(x, y)}{\partial f/\partial y(x, y)}\,.$$

Lorsque la fonction f a une forme plus compliquée, les mêmes agruments permettent de montrer que la fonction φ existe, même lorsque l'on ne peut pas la définir par une formule. On dit alors que φ est définie implicitement. Le théorème des fonctions implicites (très important en analyse et en géométrie notamment) est une généralisation de ce principe au cas où les variables x et y sont des vecteurs.

2 Énoncés des théorèmes

Théorème 1. *(thérоème d'inversion locale)*
Soient $U' \subset \mathbb{R}^n$ un ouvert, et $f\colon U' \to \mathbb{R}^n$ une application de classe $\mathcal{C}^1$. Soit $a \in U$ tel que Df_a soit inversible. Alors il exsite un voisinage ouvert $U \subset U'$ de a et un voisinage ouvert V de $f(a)$ tels que la restriction $f|_U$ de f à U soit une bijection de U sur V, de réciproque f^{-1}. En outre, $f^{-1}\colon V \to U$ est de classe $\mathcal{C}^1$, et, pour tout $x \in U$,

$$D(f^{-1})_{f(x)} = (Df_x)^{-1}\,. \tag{1}$$

Ce théorème affirme que, si la différentielle d'une fonction est inversible, alors cette fonction est *localement* une bijection. Néanmoins, même si la différentielle est partout inversible, la fonction n'est pas forcément une bijection sur son ensemble de définition.

Pour tout vecteur $v = (x, y) \in \mathbb{R}^n \times \mathbb{R}^p$, on écrit $v = (x_1, \ldots, x_n, y_1, \ldots, y_p)$ ses coordonnées dans la base canonique. Si $f\colon \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ est différentiable au point $v = (x, y)$, nous définissons $D_i f(x, y)$, $(i = 1, 2)$ par

$$D_1 f_{(x,y)} = \left(\frac{\partial f_i}{\partial x_j}(x, y)\right)_{\substack{i = 1,\ldots,p \\ j = 1,\ldots,n}} \in \mathcal{M}_{p,n}(\mathbb{R})$$
$$D_2 f_{(x,y)} = \left(\frac{\partial f_i}{\partial y_j}(x, y)\right)_{\substack{i = 1,\ldots,p \\ j = 1,\ldots,p}} \in \mathcal{M}_{p}(\mathbb{R})$$

Théorème 2. *(théorème des fonctions implicites)*
Soient $U \in \mathbb{R}^n$ et $V \in \mathbb{R}^p$ des ouverts. Soit $f\colon U \times V \to \mathbb{R}^p$ une application de classe $\mathcal{C}^1$. Soit $(a, b) \in U \times V$, tel que $f(a, b) = 0$. Supposons en outre que la matrice $D_2 f_{(a,b)}$ soit inversible. Alors, il existe un voisinage $U' \subset U$ de a, un voisinage $V' \subset V$ de b et une application $\varphi\colon U' \to V'$ tels que, pour tout $(x, y) \in U' \times V'$,

$$f(x, y) = 0 \quad \Leftrightarrow \quad y = \varphi(x)\,,$$

et tels que la matrice $D_2 f_{(x,\varphi(x))}$ soit inversible. De plus, pour tout $x \in U'$,

$$D\varphi_x = -\left(D_2 f_{(x,\varphi(x))}\right)^{-1} \cdot \left(D_1 f_{(x,\varphi(x))}\right)$$

On dit alors que l'application φ est définie implicitement par l'équation $f(x, y) = 0$ au voisinage du point (a, b).

Il existe une version plus générale de ce théorème, qui répond au problème de résoudre l'équation $f(x, y) = z$. Pour chaque z fixé, le théorème que nous avons énoncé, fournit une fonction φ_z, mais ne nous permet rien d'affirmer sur la façon dont varie la fonction φ_z en fonction de z. La généralisation en vue affirme en sus de la version ici présente que, en un sens à préciser, $z \mapsto \varphi_z$ est différentiable. L'étudiant motivé peut essayer de formuler une telle généralisation.

Ces deux théorèmes sont équivalents du point de vue logique. Cela signifie que chacun de ces théorèmes peut être démontré à partir du second. Dans ce cours, nous montrons d'abord le théorème d'inversion locale, et ensuite le thérоème des fonctions implicites à partir du précédent. Nous aurions également pu procéder dans l'autre sens.

3 Démonstration du théorème d'inversion locale

Il convient pour commencer de remarquer que si f était une application affine, alors Df_a serait constante et égale la partie linéaire de f,

$$f(x) = f(a) + Df_a(x-a)\,.$$

On pourrait alors écrire

$$x = a + (Df_a)^{-1}(f(x) - f(a))\,,$$

et le problème serait alors résolu. Mais dans le cas général, on a seulement

$$f(x) = f(a) + Df_a(x-a) + \varepsilon(x-a) \quad \text{avec } \frac{\|\varepsilon(x-a)\|}{\|x-a\|} \underset{x\to a}{\longrightarrow} 0\,.$$

On dit que ε est un petit o de $x-a$. L'objectif de la démonstration est ainsi de *chasser* ce petit o, une technique fondamentale en analyse.

Préliminaire. (d'ordre général)
Si $\|\cdot\|$ est une norme sur $\mathbb{R}^n$, alors la formule

$$\|A\| = \sup_{\|x\|\leq 1} \|Ax\|\,,$$

définit une norme sur $\mathcal{M}_n(\mathbb{R})$ qui vérifie en outre, pour toute matrice A, B,

$$\|AB\| \leq \|A\| \cdot \|B\|\,.$$

La démonstration très standard de ces propriétés est laissée en exercice. Dans la suite, l'espace vectoriel des matrices, ainsi que celui des endomorphismes de $\mathbb{R}^n$ (via la base canonique) sont munis de cette norme.
1ère étape. Montrons tout d'abord que l'on peut se ramener au cas où $a = f(a) = 0$ et $Df_a = \mathrm{Id}$ est l'application identité de $\mathbb{R}^n$ dans lui-même. La fonction $x \mapsto f(x+a) - f(a)$ vérifie les hypothèses du théorème et vaut 0 en 0. Nous pouvons donc supposer que $0 \in U'$ et $f(0) = 0$, étant donné que l'on peut retrouver la fonction f de départ à partir de cette nouvelle fonction.

Comme Df_0 est inversible, la fonction $x \mapsto (Df_0)^{-1}f(x)$ vérifie encore les hypothèses du théorème et vaut encore 0 en 0. De plus la diférentielle de cette fonction en 0 est l'identité, d'aprés la règle de dérivation d'un produit, et car la différentielle de l'application linéaire $(Df_0)^{-1}$ est elle-même. Ainsi, nous pouvons également supposer que Df_0 est l'identité pour la même raison que précédemment.

2ième étape. Montrons maintenant qu'il existe un voisinage U de 0 contenu dans U' tel que $f\colon U \to U$ est bijective.

Pour cela, considérons la fonction

$$g(x) = x - f(x)\,.$$

Cette fonction est définie sur U', elle est de classe $\mathcal{C}^1$, $g(0) = 0$, et $(Dg_0) = 0$. Comme l'application $x \mapsto Dg_x$ est continue, il existe $r > 0$ tel que

$$\text{pour tout } x \in B(O,r), \quad \|Dg_x\| < 1/2\,.$$

D'après l'inégalité des accroissements finis, nous avons alors

$$\|g(x)\| = \|g(x) - g(0)\| < 1/2\|x - 0\| = 1/2\|x\|\,,$$

ce qui montre que, si $x \in B(O,r)$, alors $g(x) \in B(O,r/2)$.

Soit maintenant $y \in B(O,r/2)$, et considérons la fonction

$$g_y(x) = y + g(x)\,.$$

Si $x \in B(O,r)$, alors $\|g_y(x)\| < r/2 + r/2 = r$, d'après le paragraphe précedent. Par suite, g_y envoie la boule $B(O,r)$ sur elle-même, et vérifie $\|D(g_y)\| = \|Dg\| < 1/2$, Toujours d'après le théorème des accroissements finis, il vient

$$\text{pour tout } x, x' \in B(O,r)\,, \quad \|g_y(x) - g_y(x')\| < 1/2\|x - x'\|\,,$$

ce qui permet d'appliquer le théorème du point fixe (théorème de Heine) à l'application g_y. Ce dernier affirme qu'il existe un unique $x_0 \in B(O,r)$, tel que $g_y(x_0) = x_0$. Ceci s'écrit également $y + g(x_0) = x_0$, et en revenant à la définition de g, nous obtenons $y = f(x_0)$. Nous avons montré que, pour tout $y \in B(O,r/2)$, il existe un unique $x \in B(O,r)$ tel que $y = f(x)$. Posons alors $U = \{x \in U'\,, f(x) \in B(O,r/2)\} \cap B(O,r)$. Nous déduisons de ce qui précède que f est une bijection de U sur son image, et que U est un voisinage de 0.

3ième étape. Nous venons de montré que f est une bijection de U sur U. Soit f^{-1} l'application réciproque. Montrons que f^{-1} est continue. Soit $y, y' \in U$, et $x = f^{-1}(y)$, $x' = f^{-1}(y')$. Il vient, en utilisant l'inégalité des accroissements finis,

$$\begin{aligned}
\|f^{-1}(y) - f^{-1}(y')\| &= \|x - x'\| \\
&= \|g(x) + f(x) - g(x') - f(x')\| \\
&\leq \|g(x) - g(x')\| + \|f(x) - f(x')\| \\
&\leq \frac{1}{2}\|x - x'\| + \|f(x) - f(x')\|
\end{aligned}$$

Ce qui s'écrit aussi

$$\|f^{-1}(y) - f^{-1}(y')\| \leq 2\|y - y'\|\,.$$

Autrement dit, f^{-1} est 2-lipschitzienne, et elle est donc continue.

4ième étape. Il reste à montrer que f^{-1} est différentiable, à donner la formule pour la dérivée, et à montrer que $x \mapsto D(f^{-1})_x$ est une application continue. Comme f est $\mathcal{C}^1$, l'application $x \to Df_x$ est continue, et comme Df_0 est l'identité, il suit que Df_x est inversible lorsque x est suffisamment proche de 0. En effet,

Lemme 3. *Supposons que $M_n(\mathbb{R})$ est munie d'une norme vérifiant, pour toute matrice A, B*

$$\|AB\| \leq \|A\|\|B\|\,.$$

Si $\|A - Id\| < 1$, alors A est inversible.

Démonstration. (esquisse) Cela marche exactement comme pour la série géométrique. La suite $X_N = \sum_{n=0}^{N}(Id - A)^n$ converge car elle est de Cauchy, et sa limite X_∞ vérifie $X_\infty A = AX_\infty = Id$. Donc A est inversible, et $A^{-1} = X_\infty$. □

Quitte à prendre U un peu plus petit, nous pouvons donc supposer que Df_x est inversible pour tout $x \in U$. Nous pouvons donc écrire, pour $y = f(x)$ et $y' = f(x')$,

$$\begin{aligned} f^{-1}(y') - f^{-1}(y) - (Df_{f^{-1}(y)})^{-1}(y - y') &= x - x' - (Df_x)^{-1})(f(x) - f(x')) \\ &= x - x' - (Df_x)^{-1}(Df_x(x - x') + \varepsilon(x - x')) \\ &= (Df_x)^{-1}\varepsilon(x - x') \end{aligned}$$

Ceci montre que f^{-1} est différentiable en y et que

$$D(f^{-1})_{f(x)} = (Df_x)^{-1} .$$

5ième étape. Il ne reste plus qu'à montrer que $y \mapsto D(f^{-1})_y$ est continue. Or, d'après la formule précédente, cela revient à montrer que l'application $A \mapsto A^{-1}$ est continue sur $\mathcal{M}_n(\mathbb{R})$, ce qui fait l'objet du lemme suivant.

Lemme 4. *Soit* $\mathrm{GL}_n(\mathbb{R})$ *l'ensemble des matrices inversibles de taille n. Alors* $\mathrm{GL}_n(\mathbb{R})$ *est un ouvert de* $\mathcal{M}_n(\mathbb{R})$*, et l'application "inverse"* $A \mapsto A^{-1}$ *est continue sur* $\mathrm{GL}_n(\mathbb{R})$.

Démonstration. $\mathrm{GL}_n(\mathbb{R})$ est l'ensemble des matrices dont le déterminant est non nul, autrement dit :

$$\mathrm{GL}_n(\mathbb{R}) = \det{}^{-1}(\mathbb{R} \setminus \{0\}) .$$

Le déterminant étant une fonction polynomiale sur $\mathcal{M}_n(\mathbb{R})$, il est continu, et en outre $\mathbb{R} \setminus \{0\}$ est un ouvert de $\mathbb{R}$. Donc $\mathrm{GL}_n(\mathbb{R})$ est l'image réciproque d'un ouvert par une application continue, et est donc un ouvert. En particulier, tous les points de $\mathrm{GL}_n(\mathbb{R})$ sont des points d'accumulation.

Pour montrer la continuité de l'inverse, utilisons la formule

$$A^{-1} = \frac{1}{\det A}\, {}^tA^{\mathrm{co}} .$$

Alors l'application "inverse" apparaît comme la composée d'applications continues, et est donc continue. En effet, $A \mapsto (\det A)^{-1}$ est continue comme nous venons de le voir. Le passage à la transposée est également continue (prendre la norme donnée par le sup). En outre, le passage à la comatrice est composée de différents déterminants, et des applications $\mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_{n-1}(\mathbb{R})$ consistant à supprimer une ligne et une colonne, ce qui est continue comme on le voit en considérant encore la norme donnée par le sup. □

4 Démonstration du théorème des fonctions implicites

La même remarque que pour le théorème d'inversion locale peut être faite. En effet, si l'on pouvait supprimer les petits o de la formule, l'équation $f(x, y) = 0$ s'écrirait

$$Ax + By = C , \quad \text{avec } A \in \mathcal{M}_{pn}(\mathbb{R}) ,\ B \in \mathcal{M}_p(\mathbb{R}) \text{ et } C \in \mathbb{R}^p ,$$

et l'hypothèse du théorème assure que B est inversible. Cela suffit pour pouvoir résoudre ce système linéaire (où y est l'inconnue) :

$$y = B^{-1}(C - Ax) = \varphi(x) .$$

Définissons la fonction $F\colon U \to \mathbb{R}^n \times \mathbb{R}^p$ en posant

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

L'application F est $\mathcal{C}^1$ et sa matrice jacobienne en (a,b) est donnée par blocs :

$$\mathrm{Jac}(F)_{(x,y)} = \begin{pmatrix} \mathrm{Id}_n & 0_{n,p} \\ D_1 f_{(x,y)} & D_2 f_{(x,y)} \end{pmatrix}.$$

Cette matrice est inversible pour (x,y) au voisinage de (a,b) car $D_2 f_{(a,b)}$ l'est par hypothèse, et son inverse est la matrice

$$\left(\mathrm{Jac}(F)_{(x,y)}\right)^{-1} = \begin{pmatrix} \mathrm{Id}_n & 0_{n,p} \\ -(D_2 f_{(x,y)})^{-1} D_1 f_{(x,y)} & (D_2 f_{(x,y)})^{-1} \end{pmatrix}.$$

Nous pouvons donc appliquer le théorème d'inversion locale à l'application F au point (a,b). Soit donc $\Omega_1 \in \mathbb{R}^n \times \mathbb{R}^p$ et $\Omega_2 \in \mathbb{R}^n \times \mathbb{R}^p$ tels que $(a,b) \in \Omega_1$, $(a, f(a,b)) = (a,0) \in \Omega_2$, et $F|_{\Omega_1}\colon \Omega_1 \to \Omega_2$ soit bijective. Soit $G\colon \Omega_2 \to \Omega_1$ son inverse. Alors $G = (G_1, G_2)$ possède pour différentielle

$$DG_{(z,w)} = \begin{pmatrix} D(G_1)_{(z,w)} \\ D(G_2)_{(z,w)} \end{pmatrix} = \begin{pmatrix} \mathrm{Id}_n & 0_{n,p} \\ -(D_2 f_{G(z,w)})^{-1} D_1 f_{G(z,w)} & (D_2 f_{G(z,w)})^{-1} \end{pmatrix}.$$

Nous en déduisons que G est de la forme $G(z,w) = (z, G_2(z,w))$. Comme $F \circ G$ est l'application identité, il vient

$$(z,w) = F(G(z,w)) = F(z, G_2(z,w)) = (z, f(z, G_2(z,w)))\,, \quad \text{et donc } w = f(z, G_2(z,w))\,.$$

L'application $\varphi(z) = G_2(z,0)$, définit sur l'ensemble U' des z tels que $(z,0) \in \Omega_2$, répond donc au problème. Pour montrer que φ est de classe $\mathcal{C}^1$, et calculer sa différentielle, il suffit de remarquer que φ est la composée

$$\begin{array}{ccccc} U' & \xrightarrow{i} & \mathbb{R}^n \times \mathbb{R}^p & \xrightarrow{G_2} & \mathbb{R}^p \\ x & \longmapsto & (x,0) & & \end{array}$$

puis d'appliquer le théorème de dérivation d'une application composée. On notera pour effectuer cette dernière vérification que $Di_x = (Id_n, 0_{n,p})$.