

Chapitre V

Calcul Différentiel

Sommaire

V.1	Dérivées partielles, notion de différentielle	79
V.1.1	Définitions, premières propriétés	79
V.1.2	Compléments	84
V.1.3	Théorème fondamental de l'analyse	89
V.1.4	Formules de changement de variable pour les transformations régulières	90
V.2	Exercices	91
V.3	Théorèmes des fonctions implicites et d'inversion locale	94
V.4	Problème adjoint	99
V.5	Exercices	102
V.6	Dérivées d'ordre supérieur	104
V.6.1	Dérivées partielles d'ordre supérieur pour les fonctions scalaires	104
V.6.2	Différentielles d'ordre supérieur pour les fonctions de \mathbb{R}^n dans \mathbb{R}^m	107
V.7	Exercices	108

V.1 Dérivées partielles, notion de différentielle

V.1.1 Définitions, premières propriétés

On sait qu'une fonction f , définie d'un intervalle ouvert $I \subset \mathbb{R}$ à valeurs dans \mathbb{R} , est dérivable en $x \in I$ si le taux de variation admet une limite, notée alors $f'(x)$, lorsque h tend vers 0, c'est-à-dire si l'on peut écrire

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \varepsilon(h),$$

où $\varepsilon(h)$ tend vers 0 quand h tend vers 0.

On a alors

$$f(x+h) = f(x) + f'(x)h + \varepsilon(h)h = f(x) + f'(x)h + o(h), \quad (\text{V.1.1})$$

où $o(h)$ (que l'on peut aussi écrire $h\varepsilon(h)$) est une fonction définie au voisinage de 0, négligeable devant $|h|$. Ce développement exprime le fait que la fonction peut être approchée à l'ordre 1 au voisinage de x par une application affine.

Inversement, si une fonction f admet au voisinage de x un développement limité du type de (V.1.1) :

$$f(x+h) = f(x) + \gamma h + o(h),$$

alors la fonction est dérivable en x , et le coefficient du terme de premier ordre est $\gamma = f'(x)$, la dérivée de f en x .

Cette approche s'étend sans difficultés au cas où la fonction est à valeurs vectorielles. Considérons

$$f : x \in \mathbb{R} \mapsto f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m.$$

Si la dérivée de chacune des composantes f_i par rapport à la variable réelle est définie en $x \in \mathbb{R}$, la dérivée $f'(x)$ s'écrit

$$f'(x) = (f'_1(x), f'_2(x), \dots, f'_m(x)),$$

et le développement limité est simplement écrit dans \mathbb{R}^m .

Nous allons nous intéresser maintenant la généralisation de ces notions au cas où l'espace de départ lui-même peut être de dimension strictement supérieure à 1, l'objet typique étudié à partir de maintenant sera donc une application

$$f : x = (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m$$

où chacune des m composantes f_i est une application de \mathbb{R}^n dans \mathbb{R} .

Exemple V.1.1. (Champ de vecteurs, champs scalaire)

Un *champ de vecteurs* dans l'espace physique est une application qui à chaque point \mathbb{R}^3 associe un vecteur de \mathbb{R}^3 . On le note en général $u = (u_1, u_2, u_3)$ où chaque composante u_i est une fonction de $x = (x_1, x_2, x_3)$. Il peut encoder un champ de vitesses fluides à un instant donné, ou un champ de déformations infinitésimales au sein d'un objet élastique déformable. Un champ de vecteurs dans le plan (par exemple un champ de vitesses horizontales à la surface d'un lac) correspond au cas $n = m = 2$.

On parle d'un *champ scalaire* lorsque l'espace d'arrivée est \mathbb{R} (cas $n = 3$ et $m = 1$ pour un champ de l'espace physique). Cela correspond par exemple au champ de température dans une zone de l'espace à un instant donné.

Exemple V.1.2. On peut considérer les versions dynamiques des exemples ci-dessus en rajoutant une variable de temps dans l'espace de départ. Par exemple un champ de vitesse variable en temps correspond au cas $n = 4, m = 3$, il peut être considéré comme une application $\mathbb{R}^3 \times \mathbb{R}$ dans \mathbb{R}^3 , qui à chaque $(x, t) = (x_1, x_2, x_3, t)$ fait correspondre un vecteur $(u_1(x, t), u_2(x, t), u_3(x, t))$.

Si l'on cherche à écrire un développement limité du type (V.1.1), l'identité est à valeurs dans \mathbb{R}^m , et la variation h de la variable x de l'espace de départ vit dans \mathbb{R}^n . Le terme $f'(x)h$ doit être remplacé par un terme à valeurs dans \mathbb{R}^m , qui dépend linéairement du vecteur $h \in \mathbb{R}^n$, il s'écrit donc sous la forme d'une application linéaire (de \mathbb{R}^n dans \mathbb{R}^m) appliquée à la variation $h \in \mathbb{R}^n$. Cette section décrit la démarche permettant d'écrire dans ce contexte multidimensionnel le développement limité d'une fonction de n variables, à valeurs dans \mathbb{R}^m , c'est à dire d'approcher localement une fonction générale par une fonction *affine*.

La notion de *différentielle*, qui généralise la notion de dérivée d'une fonction de \mathbb{R} dans \mathbb{R} , peut se définir de façon abstraite, y compris pour des espaces de dimension infinie. Nous débutons néanmoins ce chapitre par la notion plus directement accessible et utilisable de dérivée partielle, pour des fonctions de \mathbb{R}^n dans \mathbb{R}^m .

Dérivées partielles

Définition V.1.1. (Dérivées partielles, matrice jacobienne)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , et $x = (x_1, \dots, x_n) \in U$. On dit que f admet en x une dérivée partielle par rapport à la variable x_j si l'application de \mathbb{R} dans \mathbb{R}^m obtenue en figeant toutes les variables sauf la j -ième est dérivable en x_j . Plus formellement, si

$$y \mapsto f(x_1, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n) \in \mathbb{R}^m$$

définie d'un voisinage de x_j vers \mathbb{R}^m , est dérivable en $y = x_j$.

La dérivée de la i -ème composante de f par rapport à la variable x_j , telle que définie ci-dessus, est alors notée¹

$$\frac{\partial f_i}{\partial x_j}(x) \text{ ou } \partial_{x_j} f_i(x) = \lim_{s \rightarrow 0} \frac{f_i(x + s e_j) - f_i(x)}{s},$$

où l'on a noté e_j le j -ème vecteur de la base canonique de \mathbb{R}^n . Si toutes les dérivées partielles des f_i par rapport aux x_j existent, on appelle *matrice Jacobienne* la matrice

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \in \mathcal{M}_{mn}(\mathbb{R}).$$

On désignera par $\partial f / \partial x_j \in \mathbb{R}^m$ le j -ième vecteur colonne de la matrice jacobienne.

Exercice V.1.1. Écrire la matrice jacobienne (en justifiant son existence) de l'application

$$f : (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mapsto \begin{pmatrix} 2x_1 + x_2 + 3x_3 \\ x_1 x_2^2 x_3^4 \end{pmatrix} \in \mathbb{R}^2.$$

Remarque V.1.2. La définition des dérivées partielles se base sur des variations, autour du point considéré, dans les directions des axes de coordonnées, et dans ces directions seulement. Il est possible que la fonction ait un comportement pathologique si l'on considère des variations dans d'autres directions. On peut par exemple imaginer une fonction qui ne varie pas lorsque l'on perturbe selon une direction de coordonnées (on aura alors existence de dérivées partielles nulles), mais qui a un comportement très singulier dans d'autres directions (voir exercice V.1.3 ci-après). L'existence d'une matrice jacobienne (et son expression le cas échéant), n'est donc pas une propriété intrinsèque, elle dépend du système de coordonnées choisi.

Nous allons à présent définir la notion plus intrinsèque de différentiabilité d'une application de \mathbb{R}^n dans \mathbb{R}^m , dont la définition ne repose pas sur un système de coordonnées particulier. La définition repose sur l'existence d'un développement limité *uniforme* vis-à-vis de la direction de variation. Lorsque $n = 1$, nous avons rappelé précédemment que l'existence d'un développement limité est équivalente à l'existence d'une dérivée. Comme nous le verrons, cette équivalence *ne se généralise pas* au cas où l'espace de départ est de dimension ≥ 2 : l'existence de dérivées partielles en un point ne garantit pas la différentiabilité.

Notation V.1.3. (Image par une application linéaire et produit matrice vecteur)

Nous adoptons dans ce qui suit une convention courante dans le contexte du calcul différentiel (et en particulier en mécanique des fluides), qui est de noter $F \cdot x$ l'image par une application linéaire F d'un vecteur x . De la même manière, si A est une matrice, écrira $A \cdot x$ le produit matrice vecteur. Cette notation est issue de ce que l'on appelle le calcul tensoriel², qui n'est pas abordé en tant que tel dans ce cours.

Définition V.1.4. (Différentielle (•))

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . On dit que f est différentiable en $x \in U$ s'il existe une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , notée $df(x)$, telle que

$$f(x + h) = f(x) + df(x) \cdot h + o(h) \quad (\text{V.1.2})$$

où $o(h)$ est une application de \mathbb{R}^n dans \mathbb{R}^m négligeable devant $\|h\|$, c'est à dire telle que $o(h)/\|h\|$ tend vers 0 quand h tend vers 0. On appelle $df(x)$ la *différentielle* de f en x .

1. Nous commettons ici un abus de notation si courant qu'il nous paraît préférable de le commettre en connaissance de cause, plutôt que de le contourner. Dans ce qui suit x_j dans $\partial f_i / \partial x_j$ encode le fait que l'on dérive par rapport à la j -ème variable. Mais quand on écrit $x = (x_1, \dots, x_j, \dots, x_n)$, x_j désigne un réel, qui est la valeur particulière de la j -ième composante du point x considéré.

2. On pourra se reporter à http://mms2.ensmp.fr/mmc_st_etienne_fort/calcul_tensoriel/polycop/tenseurs_poly.pdf pour une présentation détaillée de ces notions.

Exercice V.1.2. a) Montrer que l'application suivante est différentiable, et préciser sa différentielle

$$f : (x_1, x_2, x_3) \in \mathbb{R}^3 \mapsto x_1 x_2^2 x_3^3$$

b) Préciser le domaine de différentiabilité de l'application

$$x \in \mathbb{R}^n \mapsto \|x\|$$

et exprimer sa différentielle lorsqu'elle existe.

Proposition V.1.5. Toute application différentiable en un point est continue en ce point.

Démonstration. C'est une conséquence directe du développement limité (V.6.2), qui assure que $f(x+h)$ tend vers $f(x)$ quand h tend vers 0. \square

Définition V.1.6. (Continue différentiabilité (•))

Une application f d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m est dite *continûment différentiable* sur U si elle est différentiable en tout $x \in U$, et si l'application $x \mapsto df(x)$ est continue sur U (l'espace d'arrivée est muni canoniquement de la norme d'opérateur subordonnée à la norme euclidienne, voir proposition I.8.6, page 28).

Lien entre différentielle et matrice jacobienne

Lorsque la différentielle existe, sa représentation dans les bases canoniques de \mathbb{R}^n et \mathbb{R}^m est la matrice jacobienne définie ci-dessus.

Proposition V.1.7. Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . Si f est différentiable en $x \in U$, alors elle admet des dérivées partielles dans toutes les directions, et sa différentielle admet pour représentation matricielle la matrice jacobienne J définie ci-dessus : c'est-à-dire que f s'écrit

$$f(x+h) = f(x) + J(x) \cdot h + o(h).$$

Démonstration. Si f est différentiable en x , on peut écrire le développement limité composante par composante, en prenant la variation h de la forme se_j , où s est un réel et e_j un vecteur unitaire de la base canonique \mathbb{R}^n : pour tout $i = 1, \dots, m$, tout $j = 1, \dots, n$,

$$f_i(x + se_j) = f_i(x) + s(df(x) \cdot e_j)_i + o(s).$$

On a donc

$$(df(x) \cdot e_j)_i = \lim_{s \rightarrow 0} \frac{f_i(x + se_j) - f_i(x)}{s} = \frac{\partial f_i}{\partial x_j}$$

d'après la définition V.1.1), c'est-à-dire le coefficient (i, j) de la matrice jacobienne J . Ce coefficient s'identifie donc à la composante i de l'image du j -ème vecteur de la base canonique par la différentielle, ce qui termine la preuve. \square

Comme nous l'avons déjà évoqué, dès que la dimension n de l'espace de départ est strictement plus grande que 1, l'existence d'une matrice jacobienne (c'est à dire l'existence de toutes les dérivées partielles) *n'implique pas* la différentiabilité. Une application peut même admettre une matrice jacobienne en un point sans pour autant être continue en ce point (voir exercice V.1.3 ci-dessous). On verra néanmoins que, si une application admet sur un ouvert U des dérivées partielles qui sont toutes *continues* sur U , alors l'application est continûment différentiable sur U (voir proposition V.1.8 ci-après).

Exercice V.1.3. Montrer que la fonction

$$(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\} \mapsto \frac{xy}{x^2 + y^2}, \quad f(0, 0) = 0,$$

admet en $(0, 0)$ des dérivées partielles, mais n'est pas continue en ce point (et donc non différentiable d'après la proposition V.1.5).

Exercice V.1.4. (●) Montrer que toute application affine définie d'un ouvert de \mathbb{R}^n dans \mathbb{R}^m :

$$x \mapsto f(x) = A \cdot x + b, \quad A \in \mathcal{M}_{mn}(\mathbb{R}), \quad b \in \mathbb{R}^m,$$

est continûment différentiable sur cet ouvert, et préciser sa différentielle.

Exercice V.1.5. (●) a) Soient f et g deux applications différentiables sur un ouvert $U \subset \mathbb{R}^2$, à valeurs dans \mathbb{R} . Exprimer la différentielle de l'application produit

$$F : (x_1, x_2) \in U \mapsto f(x_1, x_2)g(x_1, x_2).$$

b) Soient f (respectivement g) une application différentiable sur un ouvert U (respectivement V) de \mathbb{R}^2 dans \mathbb{R} . Exprimer la différentielle de l'application

$$G : (x_1, x_2, x_3, x_4) \in U \times V \mapsto f(x_1, x_2)g(x_3, x_4),$$

et écrire sa jacobienne en fonction de celles de f et g . On pourra utiliser la notation $x_{12} = (x_1, x_2) \in \mathbb{R}^2$, $h_{12} = (h_1, h_2) \in \mathbb{R}^2$, et de même pour les indices 3 et 4.

D'après la proposition V.1.7, si une application est continûment différentiable sur un ouvert U , alors la matrice jacobienne est définie en tout point de cet ouvert, et la correspondance $x \mapsto J(x)$ est continue. La proposition suivante assure la réciproque de cette propriété.

Proposition V.1.8. (●●) Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . On suppose que la matrice jacobienne $J(x)$ est définie en chaque point x de U , et que l'application $x \mapsto J(x)$ est continue. Alors f est continûment différentiable sur U .

Démonstration. On écrit la démonstration pour le cas $n = 2$, et l'on suppose que $(0, 0) \in U$ pour simplifier les notations. Nous allons montrer la différentiabilité en $(0, 0)$, la démonstration pour les autres points étant essentiellement la même. Pour h_1, h_2 suffisamment petits, on a

$$f(h_1, h_2) - f(0, 0) = f(h_1, h_2) - f(h_1, 0) + f(h_1, 0) - f(0, 0) + f(0, 0) - f(0, 0).$$

D'après le théorème des accroissements finis, il existe $\theta_1 \in]0, 1[$ (qui dépend de h_1) tel que

$$f(h_1, 0) - f(0, 0) = h_1 \partial_1 f(\theta_1 h_1, 0).$$

Comme $x \mapsto J(x)$ est continue, tous les coefficients de la matrice J sont des fonctions continues en x . On a donc en particulier³

$$\partial_1 f(\theta_1 h_1, 0) = \partial_1 f(0, 0) + \varepsilon(h_1),$$

et ainsi

$$f(h_1, 0) - f(0, 0) = h_1 \partial_1 f(0, 0) + h_1 \varepsilon(h_1).$$

Il existe de la même manière $\theta_2 \in]0, 1[$ (qui dépend de h_1 et de h_2)

$$f(h_1, h_2) - f(h_1, 0) = h_2 \partial_2 f(h_1, \theta_2 h_2),$$

avec, par continuité des dérivées partielles⁴,

$$\partial_2 f(h_1, \theta_2 h_2) = \partial_2 f(0, 0) + \varepsilon(h).$$

3. Le reste négligeable s'écrit a priori $\varepsilon(\theta_1 h_1)$, mais comme θ_1 , qui dépend de h_1 , est toujours dans $]0, 1[$, il s'agit bien d'une quantité qui tend vers 0 quand h_1 tend vers 0.

4. Comme précédemment, le reste s'écrit a priori $\varepsilon(h_1, \theta_2 h_2)$, avec $\theta_2 \in]0, 1[$ qui dépend de $h = (h_1, h_2)$, il s'agit donc bien d'une fonction de h qui tend vers 0 quand $h = (h_1, h_2)$ tend vers $(0, 0)$.

On a donc

$$f(h_1, h_2) - f(h_1, 0) = h_2 \partial_2 f(0, 0) + h_2 \varepsilon(h).$$

On a donc finalement

$$f(h_1, h_2) = f(0, 0) + h_1 \partial_1 f(0, 0) + h_2 \partial_2 f(0, 0) + o(h),$$

qui exprime la différentiabilité de f en $(0, 0)$, et de la même manière en tout point de l'ouvert U . La différentielle peut s'exprimer matriciellement à partir de la jacobienne $J = [\partial_1 f, \partial_2 f]$ (écriture de la matrice en colonnes, chacune des dérivées partielles étant un vecteur de \mathbb{R}^m). Les dérivées partielles étant continues, la correspondance $x \mapsto J(x)$ est continue, l'application est donc continûment différentiable sur U .

La réciproque est une conséquence directe de la proposition V.1.7. \square

Proposition V.1.9. (Différentielle de la composée de deux applications)

Soient g une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^p , et f définie d'un ouvert $V \subset \mathbb{R}^p$ dans \mathbb{R}^m . On suppose que $g(U) \subset V$. Si g est différentiable en $x \in U$ et f est différentiable en $g(x)$, alors $f \circ g$ est différentiable en x , et l'on a

$$d(f \circ g)(x) = df(g(x)) \circ dg(x).$$

Démonstration. On a

$$\begin{aligned} f \circ g(x + h) - f \circ g(x) &= f(g(x) + dg(x) \cdot h + o(h)) \\ &= f(g(x)) + df(g(x)) \cdot (dg(x) \cdot h + o(h)) + o(dg(x) \cdot h + o(h)) \\ &= f \circ g(x) + (df(g(x)) \circ dg(x)) \cdot h + o(h), \end{aligned}$$

qui exprime la différentiabilité de $f \circ g$, avec l'expression annoncée de la différentielle. \square

Exercice V.1.6. a) Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$, $a \in \mathbb{R}^n$, et f une application différentiable de \mathbb{R}^n dans \mathbb{R}^m . Exprimer la différentielle et la matrice jacobienne de $F : x \mapsto f(b + Ax)$.

b) Soit f une fonction dérivable de \mathbb{R} dans \mathbb{R} . Déterminer la différentielle de

$$F : (x, y) \in \mathbb{R}^2 \mapsto f(x^2 + y^2),$$

et écrire le développement limité de F au voisinage d'un point $(x, y) : F(x + h_x, y + h_y) = \dots$

Remarque V.1.10. (Et en pratique, on fait comment ?)

Les définitions ci-dessus suggèrent deux stratégies pour étudier la différentiabilité d'une application donnée, et préciser sa différentielle : calculer sa matrice jacobienne J_f , et identifier le domaine sur lequel elle est définie et continue, ou effectuer un développement limité $J(f + h)$ et en extraire la partie d'ordre 1 en h . Chacune de ses approches peut se révéler la plus pertinente dans certains cas. Ainsi pour l'application de l'exercice V.1.1, il est plus aisé d'écrire directement la matrice jacobienne que d'effectuer le développement limité. Pour la seconde application de l'exercice V.1.2 en revanche, le développement limité permet d'arriver de façon plus directe et élégante à l'expression de la différentielle. Noter qu'en dimension infinie la notion de différentielle peut se définir comme nous l'avons fait en dimension finie (voir définition III.1.18, page 61), alors que la notion de matrice jacobienne n'a a priori pas de sens. L'approche par développement limité est de ce point de vue plus générale, et plus universelle puisqu'elle ne dépend pas des bases des espaces de départ ou d'arrivée.

V.1.2 Compléments

Proposition V.1.11. (Linéarité de la différentiation)

Soient f et g des applications définies d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . Si f et g sont différentiables en $x \in U$, alors, pour tous λ, μ réels, l'application $\lambda f + \mu g$ est différentiable, et l'on a

$$d(\lambda f + \mu g) = \lambda df + \mu dg.$$

Démonstration. C'est une conséquence immédiate de la définition de la différentiabilité. \square

Corollaire V.1.12. L'ensemble $C^1(U, \mathbb{R}^m)$ des applications continûment différentiables sur un ouvert U de \mathbb{R}^n est un espace vectoriel.

Notion de gradient

Lorsqu'une fonction différentiable est à valeurs dans \mathbb{R} , la différentielle est une *forme linéaire*, c'est-à-dire une application linéaire de \mathbb{R}^n dans \mathbb{R} . Elle peut alors s'exprimer⁵ à l'aide du produit scalaire canonique sur \mathbb{R}^n , et s'identifie par ce biais à un vecteur de \mathbb{R}^n . C'est ce vecteur que l'on appelle gradient de f .

Définition V.1.13. (Gradient)

Soit f une application définie d'un ouvert U de l'espace euclidien \mathbb{R}^n , à valeurs dans \mathbb{R} . On note $\langle \cdot | \cdot \rangle$ le produit scalaire sous-jacent (qui peut être, ou ne pas être, le produit scalaire canonique). On suppose que f est différentiable en $x \in U$. Il existe alors un unique vecteur, noté $\nabla f(x)$, tel que

$$df(x) \cdot h = \langle \nabla f(x) | h \rangle \quad \forall h \in \mathbb{R}^n.$$

Conformément à la proposition V.1.7, si le gradient est relatif au produit scalaire canonique, il s'écrit, dans la base canonique,

$$\nabla f(x) = \begin{pmatrix} \partial_{x_1} f(x) \\ \partial_{x_2} f(x) \\ \vdots \\ \partial_{x_n} f(x) \end{pmatrix} \in \mathbb{R}^n. \quad (\text{V.1.3})$$

Remarque V.1.14. Le gradient d'un champ scalaire est présenté dans certains ouvrages comme le vecteur dont les composantes sont les dérivées partielles selon les composantes de l'espace de départ (équation (V.1.3) ci-dessus). Cette définition laisse entendre qu'il suffit que ces dérivées partielles existent pour que le gradient existe, et aucun lien n'est fait avec un quelconque produit scalaire. On prendra garde au fait que la définition rigoureuse exige que la fonction soit différentiable au point considéré, ce qui est strictement plus fort que l'existence de dérivées partielles (voir exercice V.1.3, page 82). Par ailleurs la notion de gradient est intimement liée à la structure euclidienne (ou hilbertienne en dimension infinie) de l'espace, elle est le produit de l'identification d'une forme linéaire avec un vecteur, au travers d'un produit scalaire (théorème de Riesz-Fréchet III.1.17, page 60, pour les espace hilbertiens de dimension finie ou infinie), de sorte que ce gradient *dépend du produit scalaire choisi* (voir exercices V.1.8, V.1.9, et V.1.10 ci-après).

Exercice V.1.7. Reprendre l'exercice V.1.6 (en supposant $m = 1$ pour la première question), en précisant dans chaque cas le gradient de F en fonction de celui de f .

Il sera utile dans certaines applications d'utiliser la notion de *gradient partiel*. Cela consiste simplement à considérer une fonction de n variables comme une fonction d'une partie de ces variables, les autres étant gelées. La définition ci-dessous précise cette notion dans le cas général, que nous illustrons au préalable sur un cas particulier. Considérons une fonction de \mathbb{R}^3 dans \mathbb{R} , différentiable en un point $x = (x_1, x_2, x_3)$. Son gradient en x est un vecteur de \mathbb{R}^3 . Si on la considère maintenant comme une fonction de (x_1, x_2) , avec x_3 fixé à sa valeur correspondant à x , le gradient de cette nouvelle fonction est un vecteur de \mathbb{R}^2 , que l'on pourra noter $\nabla_{x_1 x_2} f$, ou $\nabla_{x_{12}} f$.

Définition V.1.15. (Gradient partiel)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , différentiable en un point $x \in U$. On écrit

5. Lorsque l'on travaille sur \mathbb{R}^n , l'usage de la base orthonormée canonique et du produit scalaire canonique sont tellement naturels que l'on a tendance à identifier spontanément la différentielle et le gradient. On prendra cependant garde au fait que le gradient n'est pas défini de façon intrinsèque. Contrairement à la différentielle, qui est une application définie de façon non ambiguë par le développement limité, ce gradient dépend du produit scalaire choisi. Ce point est particulièrement sensible dans certaines situations, notamment en dimension infinie, où plusieurs produits scalaires "naturels" peuvent co-exister.

$\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_p}$, de telle sorte que x peut s'écrire⁶ $x = (x_1, \dots, x_p)$, avec $x_j \in \mathbb{R}^{n_j}$. Pour i entre 1 et p , on considère la fonction partielle qui ne dépend que du vecteur x_j , les autres étant fixées. Le gradient de cette fonction partielle est un vecteur de \mathbb{R}^{n_j} , on le note $\nabla_{x_j} f$.

Cette notion de gradient partiel est notamment très utile en pratique lorsque l'on définit un potentiel d'interaction sur un système de particules localisées en q_1, q_2, \dots, q_N , chacun des q_i étant un point de l'espace physique \mathbb{R}^3 . Si l'on définit un potentiel d'interaction sur le système comme une fonction Φ de $q = (q_1, \dots, q_N) \in \mathbb{R}^{3N}$, la force exercée sur la particule i dérivant de ce potentiel d'interaction est simplement $-\nabla_{q_i} \Phi \in \mathbb{R}^3$. On se reportera à l'exercice V.2.8, page 93, pour une étude plus approfondie de ces systèmes de particules en interaction, et l'utilisation dans ce cadre de la notion de gradient partiel.

Remarque V.1.16. Si l'on considère le gradient partiel d'une fonction de $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ vis-à-vis d'une unique variable x_j , on retrouve la notion de dérivée partielle par rapport à x_j déjà introduite.

Définition V.1.17. (Différentielle partielle)

On définit de façon tout à fait analogue une notion de différentielle partielle, pour des applications de \mathbb{R}^n vers \mathbb{R}^m . Si l'on écrit comme précédemment

$$x = (x_1, \dots, x_p) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_p},$$

la différentielle partielle par rapport à x_j , que l'on notera⁷ $\partial_{x_j} f$, est alors une application linéaire de \mathbb{R}^{n_j} dans \mathbb{R}^m .

Exercice V.1.8. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive, de telle sorte que

$$(x, y) \mapsto \langle x | y \rangle_A = \langle Ax | y \rangle$$

puisse être utilisé comme produit scalaire sur \mathbb{R}^d . On note ∇_A l'opérateur gradient pour ce produit scalaire. Pour J fonctionnelle continûment différentiable, exprimer $\nabla_A J$ en fonction de ∇J

Démonstration. On a

$$J(x+h) = J(x) + dJ(x) \cdot h + o(h) = J(x) + \langle \nabla J(x) | h \rangle + o(h) =$$

$$J(x) + \langle AA^{-1} \nabla J(x) | h \rangle + o(h) = J(x) + \langle A^{-1} \nabla J(x) | h \rangle_A + o(h),$$

d'où $\nabla_A J = A^{-1} \nabla J$. □

Exercice V.1.9. (Dépendance du gradient vis-à-vis du produit scalaire, système de particules)

Comme indiqué précédemment, le gradient dépend du produit scalaire sous-jacent. Dans le cadre de ce cours, nous utiliserons cette notion essentiellement en lien avec le produit scalaire canonique de \mathbb{R}^n , parfois sans re-préciser qu'il s'agit bien de ce produit scalaire. Cette exercice illustre le fait qu'il peut être naturel, dans certains contextes, de travailler avec d'autres produits scalaires, et permet de comprendre comment le gradient se voit modifié.

On se place dans $\mathbb{R}^{3N} = \mathbb{R}^3 \times \cdots \times \mathbb{R}^3$ pour représenter les vitesses dans l'espace physique de N particules, de masses m_1, m_2, \dots, m_N toutes strictement positives. On considère la fonction qui à un jeu de vitesses associe l'énergie cinétique

$$E(u) = E(u_1, \dots, u_N) = \frac{1}{2} \sum_{n=1}^N m_n |u_n|^2.$$

6. Attention, x_j désigne dans ce qui suit non plus une variable scalaire, mais un groupe de n_j variables scalaires.

7. Cette notation est la plus couramment utilisée, et nous en recommandons l'usage, tout en reconnaissant qu'il aurait été assez naturel d'utiliser la notation d_{x_j} , puisqu'il s'agit d'une application linéaire, définie de façon intrinsèque comme associant à tout vecteur un vecteur, indépendamment du choix d'une base. La notation ' ∂ ' a pour l'instant été utilisée pour représenter des dérivées partielles, qui reposent sur le choix d'un système de coordonnées. Dans le cas présent, on est un peu entre les deux : on souhaite représenter une application linéaire, mais définie sur un sous-espace dont la définition repose sur le choix d'un système de coordonnées.

Montrer que E est différentiable, et calculer son gradient pour le produit scalaire canonique, puis pour le produit scalaire $\langle \cdot | \cdot \rangle_m$ pondéré par la collection de masses $m = (m_1, \dots, m_N)$, défini par

$$\langle u | v \rangle_m = \sum_{n=1}^N m_n \langle u_n | v_n \rangle,$$

où $\langle u_n | v_n \rangle$ représente le produit scalaire canonique sur \mathbb{R}^3 .

Exercice V.1.10. Soient a_1 et a_2 deux réels strictement positifs. On considère la forme bilinéaire

$$(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto (x, x) = \frac{x_1 y_1}{a_1^2} + \frac{x_2 y_2}{a_2^2},$$

et l'on note $J(x) = a(x, x)/2$.

Montrer que cette forme $a(\cdot, \cdot)$ définit un produit scalaire sur \mathbb{R}^2 . Préciser la sphère unité pour ce produit scalaire. On considère un point $x = (x_1, x_2)$ de cette sphère unité S , n le gradient de J en x (selon le produit scalaire canonique), et n_a le gradient de J en x selon le produit scalaire défini par a . Décrire les droites qui passent par x et qui ont pour directions n et n_a .

Que ce passe-t-il en dimension $d > 2$?

Définition V.1.18. (Point critique / stationnaire)

Soit f une application continûment différentiable d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} . On appelle point *critique* ou point *stationnaire* tout point $x \in U$ en lequel le gradient s'annule.

Exercice V.1.11. (•) Justifier l'appellation *stationnaire* dans la définition précédente.

Calcul différentiel

Nous regroupons ici quelques considérations sur la pratique effective du calcul différentiel, et en particulier les notations dx_1 , $dx_1 + dx_2$, etc ...

Si l'on considère par exemple une fonction de \mathbb{R}^2 dans \mathbb{R} , définie par $f(x) = f(x_1, x_2) = x_1^2 + x_2^3 + x_1 x_2$, on écrira

$$df = d(x_1^2 + x_2^3 + x_1 x_2) = 2x_1 dx_1 + 3x_2 dx_2 + x_1 dx_2 + x_2 dx_1 = (2x_1 + x_2) dx_1 + (3x_2 + x_1) dx_2.$$

Dans ce qui précède, dx_1 représente par exemple la différentielle de la fonction $(x_1, x_2) \mapsto x_1$, qui est simplement l'application qui à (h_1, h_2) associe h_1 , qui peut se représenter matriciellement par $[1 \ 0]$. La différentielle de f en (x_1, x_2) est donc représentée dans la base canonique par

$$J(x) = [2x_1 + x_2 \quad 3x_2 + x_1].$$

De façon plus générale, on écrira⁸

$$df(x_1, x_2) = \frac{\partial f}{\partial x_1}(x_1, x_2) dx_1 + \frac{\partial f}{\partial x_2}(x_1, x_2) dx_2.$$

On prendra garde au fait que l'expression dx_1 dépend du contexte. La même expression peut correspondre par exemple à une application de \mathbb{R}^3 dans \mathbb{R} , auquel cas dx_1 est représentée matriciellement par $[1 \ 0 \ 0]$.

8. Nous commettons ici un abus de notation courant, auquel il convient de s'habituer car il est très répandu : le ' x_1 ' qui apparaît dans ∂_{x_1} et dans dx_1 représente une variable générique vis à vis de laquelle on dérive, alors que le ' x_1 ' de $df(x_1, x_2)$ est un nombre réel, première coordonnée du point en lequel on dérive la fonction. On devrait en toute rigueur distinguer ces deux acceptions en utilisant des noms différents, par exemple

$$df(a_1, a_2) = \partial_{x_1} f(a_1, a_2) dx_1 + \partial_{x_2} f(a_1, a_2) dx_2.$$

Si l'application est à valeurs vectorielles, par exemple dans \mathbb{R}^2 :

$$f(x_1, x_2) = \begin{bmatrix} x_1^2 + x_2^3 + x_1 x_2 \\ x_1 \end{bmatrix},$$

on écrira de la même manière

$$df(x_1, x_2) = \begin{bmatrix} (2x_1 + x_2) dx_1 + (3x_2 + x_1) dx_2 \\ dx_1 \end{bmatrix},$$

qui peut se représenter matriciellement par

$$J(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 & 3x_2 + x_1 \\ 1 & 0 \end{bmatrix},$$

de telle sorte que, pour tout h dans \mathbb{R}^2 , on a le développement limité

$$f(x + h) = f(x) + J(x) \cdot h + o(h),$$

où $J(x) \cdot h$ représente le produit matrice vecteur, comme indiqué précédemment.

Exercice V.1.12. Calculer la différentielle de la forme de Minkovski

$$f : (x, y, z, t) \in \mathbb{R}^4 \longmapsto x^2 + y^2 + z^2 - c^2 t^2,$$

avec $c > 0$ (vitesse de la lumière).

Récapitulatif

Les développements ci-dessus décrivent des manières variées d'exprimer qu'une fonction peut être approchée localement par une fonction affine. Nous récapitulons ici ces différentes manières, en rappelant leur cadre d'utilisation et les liens entre elles. Dans ce qui suit f désigne, sauf mention contraire, une application de \mathbb{R}^n dans \mathbb{R}^m .

Comme on l'a vu, f est dite différentiable en $x \in \mathbb{R}^n$ s'il existe une application $df(x)$ telle que

$$f(x + h) = f(x) + df(x) \cdot h + o(h).$$

Le terme $df(x) \cdot h$ désigne l'image par $df(x)$ du vecteur h . Cette expression est intrinsèque, au sens où elle ne dépend pas du choix d'une base. En pratique, on assimile souvent une application linéaire et son écriture matricielle dans la base canonique, mais il est important de garder en tête la différence entre les deux. Cette approche permet notamment une extension immédiate de la définition en dimension infinie, dans un contexte où les bases sont inutilisables.

Si f est différentiable en x , alors (proposition V.1.7) la différentielle admet une représentation matricielle dans la base canonique qui est la matrice jacobienne $J = (\partial_{x_j} f_i)$. On a donc

$$f(x + h) = f(x) + J(x) \cdot h + o(h),$$

où $J(x) \cdot h$ est maintenant un produit matrice-vecteur. L'objet $J(x)$ dépend du choix de la base. L'expression ci-dessus peut être détaillée, de différentes manières. On peut l'écrire composante par composante

$$f_i(x + h) = f_i(x) + \sum_{j=1}^m \partial_{x_j} f_i(x) h_j + o(h),$$

ou de façon globale, avec e_i le i -ème vecteur de la base canonique de \mathbb{R}^m :

$$f(x + h) = f(x) + \sum_{i=1}^n \sum_{j=1}^m \partial_{x_j} f_i(x) h_j e_i + o(h),$$

Comme il a été précisé, l'existence de dérivées partielles en un point ne garantit pas la différentiabilité. En revanche (proposition V.1.8), si les dérivées partielles sont définies et continues sur un ouvert, alors la fonction est continûment différentiable sur cet ouvert.

Lorsque la fonction est à valeurs dans \mathbb{R} (cas $m = 1$), la matrice jacobienne est une matrice ligne, et l'application différentielle $df(x)$ est une forme linéaire. On peut alors écrire $df(x) \cdot h$ sous la forme d'un produit scalaire $\langle g | h \rangle$, où g est appelé *gradient* de f au point x , et noté $\nabla f(x)$. On a alors le développement

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + o(h).$$

Le vecteur $\nabla f(x)$ dépend du produit scalaire choisi. Lorsque ce choix n'est pas précisé, il s'agit du gradient associé au produit scalaire canonique sur l'espace euclidien \mathbb{R}^n . Lorsque l'on se place dans la base canonique de \mathbb{R}^n , que l'on considère muni du produit scalaire canonique, le vecteur $\nabla f(x)$ est représenté dans la base canonique par la matrice-ligne $J(x)$:

$$\nabla f(x) = (\partial_{x_1} f, \partial_{x_2} f, \dots, \partial_{x_n} f).$$

V.1.3 Théorème fondamental de l'analyse

Le théorème fondamental de l'analyse peut prendre plusieurs formes selon le sens que l'on donne à la notion d'intégrale. Le chapitre sur l'intégrale de Lebesgue montre que l'on peut définir cette intégrale pour des classes très générales de fonctions. Nous nous en limiterons ici à une définition plus classique de l'intégrale, en nous limitant à des fonctions continûment différentiables, de telle sorte que l'on n'aura besoin d'intégrer que des fonctions continues. On pourra donc s'en tenir à la notion d'intégrale de Riemann. L'objet de cette section est de généraliser au cas vectoriel la propriété portant sur les fonctions de \mathbb{R} dans \mathbb{R}^m : pour toute fonction f continûment dérivable sur $]a, b[$, à valeurs dans \mathbb{R}^m , pour tout x dans $]a, b[$, tout h tel que $x + h \in]a, b[$, on a

$$f(x + h) = f(x) + \int_x^{x+h} f'(s) ds.$$

Cette intégrale peut s'écrire différemment en introduisant la fonction $t \in [0, 1] \mapsto f(x + th)$, donc la dérivée en t est $f'(x + th)h$. On a

$$f(x + h) = f(x) + \int_0^1 f'(x + th)h dt.$$

Le théorème suivant généralise cette propriété aux fonctions de plusieurs variables.

Théorème V.1.19. Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , continûment différentiable sur U , et h tel que le segment

$$[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$$

soit inclus dans U . On a alors

$$f(x + h) = f(x) + \int_0^1 df(x + th) \cdot h dt.$$

Démonstration. On introduit l'application Φ de $[0, 1]$ dans \mathbb{R}^m , définie par

$$\Phi : t \in [0, 1] \mapsto \Phi(t) = f(x + th).$$

D'après la proposition V.1.9, cette application est continûment différentiable (on dira plus simplement *dérivable*, puisqu'il s'agit d'une fonction de \mathbb{R} dans \mathbb{R}^m), de dérivée

$$\Phi'(t) = df(x + th) \cdot h.$$

On a donc

$$f(x + h) - f(x) = \Phi(1) - \Phi(0) = \int_0^1 \Phi'(t) dt = \int_0^1 df(x + th) \cdot h dt,$$

qui est l'identité annoncée. □

Ce théorème nous conduit naturellement au théorème des accroissements finis pour les fonctions de plusieurs variables, qui exprime un principe simple que l'on retrouve dans différents contextes⁹ : si l'on contrôle les variations d'une certaine quantité le long d'un chemin de longueur finie qui va de x vers $x + h$, alors on peut contrôler la différence des valeurs entre $x + h$ et x .

Théorème V.1.20. (des accroissements finis)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , continûment différentiable sur U , et h tel que le segment

$$[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$$

soit inclus dans U . Alors

$$\|f(x + h) - f(x)\| \leq \max_{t \in [0, 1]} \|df(x + th)\| \|h\|,$$

où $\|df(x + th)\|$ est la norme de l'application linéaire $df(x + th)$ de \mathbb{R}^n dans \mathbb{R}^m (norme subordonnée à la norme euclidienne, selon la définition I.8.6, page 28).

Démonstration. Notons en premier lieu que, la différentielle df étant continue sur le compact $[x, x + h]$, elle est bien bornée et atteint ses bornes, en particulier le max ci-dessus est bien défini comme un réel positif. On prend la norme de l'identité établie dans le théorème précédent : On a alors

$$\begin{aligned} \|f(x + h) - f(x)\| &= \left\| \int_0^1 df(x + th) \cdot h \, dt \right\| \leq \int_0^1 \|df(x + th) \cdot h\| \, dt \\ &\leq \int_0^1 \|df(x + th)\| \|h\| \, dt \leq \max_{t \in [0, 1]} \|df(x + th)\| \|h\|, \end{aligned}$$

qui est bien l'inégalité annoncée. □

Exercice V.1.13. L'inégalité établie précédemment est-elle valide si l'on munit \mathbb{R}^n et \mathbb{R}^m d'autres normes que la norme euclidienne ?

Exercice V.1.14. Soit f une application continûment différentiable sur un ouvert U , et $K \subset U$ un compact convexe. Montrer que f est Lipchstizienne sur K .

V.1.4 Formules de changement de variable pour les transformations régulières

Nous énonçons ici sans démonstration les formules de changement de variables pour une transformation régulière.

Proposition V.1.21. Soient U un ouvert de \mathbb{R}^d et T un C^1 -difféomorphisme (définition V.3.7, page 98) entre U et $V \subset \mathbb{R}^d$.

(i) Pour tout borélien B de U

$$\lambda(T(B)) = \int_B |\det J_T| \, d\lambda,$$

où $|\det J_T|$ est la valeur absolue du déterminant de la matrice jacobienne de T .

(ii) Pour toute fonction f de V dans $\overline{\mathbb{R}}$, mesurable, alors f est intégrable sur V si et seulement si $|\det J_T| f \circ T$ est intégrable sur U , et l'on a alors, sur tout borélien dans U ,

$$\int_{T(B)} f(y) \, d\lambda(y) = \int_B f(T(x)) |\det J_T| \, d\lambda.$$

9. On pourra penser à une version *Tour de France* de cette propriété très générale : si un coureur cycliste effectue un parcours de 10 km sur une route dont la pente n'excède pas 7 %, il sait qu'il n'aura pas monté en altitude de plus de $10 \text{ km} \times 0.07 = 700 \text{ m}$.

V.2 Exercices

Exercice V.2.1. Soit f la fonction de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(x_1, x_2) = \max(x_1, x_2).$$

- a) Montrer que f est continue sur \mathbb{R}^2 .
- b) Montrer que f est continûment différentiable sur

$$\{(x_1, x_2) \in \mathbb{R}^2, x_1 \neq x_2\},$$

et préciser sa différentielle et son gradient sur chaque composante de cet ensemble (de part et d'autre de la diagonale).

- c) Montrer que f n'est pas différentiable sur $\{(x, x), x \in \mathbb{R}\}$.

Exercice V.2.2. (••) Soit $A \in \mathcal{M}(\mathbb{R}^{n \times n})$ une matrice carrée.

- a) Montrer que la fonction

$$f : x \mapsto f(x) = \frac{1}{2} \langle Ax | x \rangle - \langle b | x \rangle \in \mathbb{R}, \quad b \in \mathbb{R}^n,$$

est différentiable, et préciser son gradient.

- b) Quelle forme prend ce gradient si A est symétrique ?
- c) Quels sont les points stationnaires de f ?

Exercice V.2.3. (Vecteur gaussien)

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée, que l'on suppose *symétrique définie positive*, c'est-à-dire que $A^T = A$, et $\langle Ax | x \rangle > 0$ pour tout $x \neq 0$. On s'intéresse à la fonction qui représente (à constante de normalisation près) la loi d'un vecteur gaussien centré en a dans \mathbb{R}^n :

$$f(x) = \exp \left(-\frac{1}{2} \langle A \cdot (x - a) | x - a \rangle \right).$$

- a) Montrer que f est différentiable, et donner l'expression de son gradient.
- b) Quels sont les points stationnaires de f ?

Exercice V.2.4. (Fonctions holomorphes)

Soit Ω un ouvert de \mathbb{C} . On dit qu'une fonction f de Ω dans \mathbb{C} est *holomorphe* sur Ω si, pour tout z de Ω , la quantité

$$\frac{f(z+h) - f(z)}{h}$$

admet une limite quand $h \in \mathbb{C}$ tend vers 0. On note alors $f'(z)$ cette valeur.

Montrer l'équivalence entre les deux assertions

- (i) La fonction f est dérivable sur Ω , et l'application $z \mapsto f'(z)$ est continue.
- (ii) La fonction f vue comme application de \mathbb{R}^2 dans \mathbb{R}^2 est continûment différentiable, et l'on a sur Ω

$$\partial_x f + i \partial_y f = 0. \tag{V.2.1}$$

Exercice V.2.5. (••) On se place sur \mathbb{R}^2 muni de la distance euclidienne. Pour un ensemble donné du plan $A \subset \mathbb{R}^2$, on considère la fonction définie par $f(x) = d(x, A)$ (distance du point x à l'ensemble A).

- a) Préciser les zones de différentiabilité de f lorsque A est (i) un singleton, (ii) une paire de points distincts, (iii) un cercle, (iv) un disque, (v) un rectangle, (vi) une forme "quelconque"...
- b)(*) On associe à chaque grande ville de France (pour fixer les idées on pourra imaginer les 20 plus grandes villes par exemple) un point (son barycentre), et l'on appelle A l'ensemble de ces points. Que peut-on dire des points de non différentiabilité de la fonction f définie ci-dessus ?

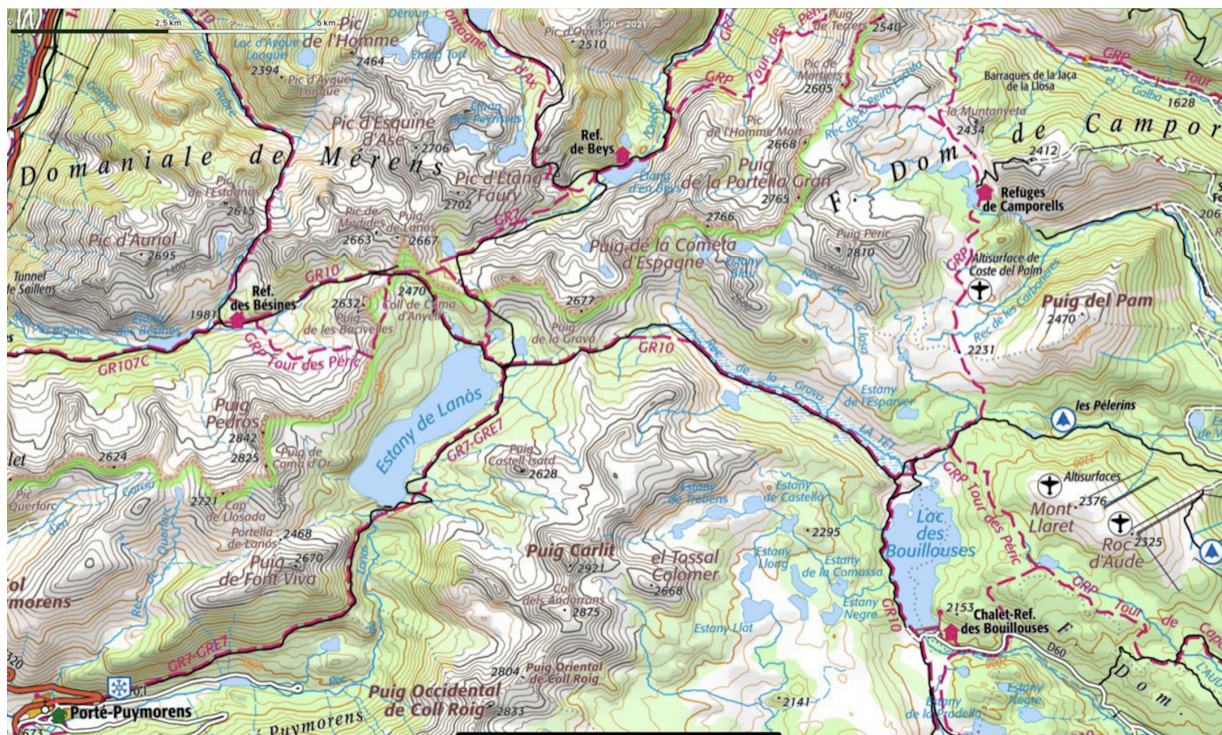


FIGURE V.2.1 – Isovaleurs de l'altitude sur une zone des Pyrénées

Exercice V.2.6. La figure V.2.1 représente les isovaleurs de la fonction altitude pour une certaine zone géographique, que l'on peut considérer comme une fonction de \mathbb{R}^2 dans \mathbb{R} .

- Localiser des points critiques de cette fonction f (c'est à dire le point en lesquels le gradient s'annule), et décrire la forme de la fonction f au voisinage de ces points. Proposer des fonctions polynomiales qui vous paraissent de nature à reproduire la forme de la fonction au voisinage du point critique, dans les différents cas.
- (★) Comment caractériser les zones correspondant aux lacs ?
- (★) Comment peut-on caractériser le bassin d'attraction d'un lac, c'est à dire l'ensemble des x tels qu'une goutte d'eau tombée en x va alimenter le lac en question ?
- (★) Le nombre de lacs peut-il augmenter ou diminuer en fonction de la pluviométrie ?

Exercice V.2.7. (Taux de déformation (●●))

On considère un champ de vitesses (on pourra penser à la vitesse instantanée d'un fluide occupant une partie de l'espace physique)

$$x = (x_1, x_2, x_3) \mapsto u(x) = (u_1(x), u_2(x), u_3(x)).$$

On suppose u différentiable en un point x d'un ouvert U de \mathbb{R}^3 .

- Montrer que l'on peut écrire le champ de vitesse au point $x + h$ voisin de x de la façon suivante

$$u(x + h) = u(x) + \omega \wedge h + D \cdot h + o(h), \quad (\text{V.2.2})$$

où ω est un vecteur de \mathbb{R}^3 , et $D \in \mathcal{M}_3(\mathbb{R})$ une matrice symétrique.

- Justifier l'appellation *matrice des taux de déformation* utilisée pour désigner la matrice D .
- (★) On dit qu'un écoulement est incompressible si $\partial_1 u_1 + \partial_2 u_2 + \partial_3 u_3 = 0$ (i.e. si ce qu'on appelle la *divergence* de u est nulle). Montrer que si l'écoulement est incompressible sur U , alors la somme des valeurs

propres de la matrice D associée à tout point de U est égale à 0, et interprétez physiquement cette propriété.

d) Donner un exemple de champ de vitesses u non trivial défini sur \mathbb{R}^3 tel qu'en tout point, la décomposition (V.2.2) soit telle que $D = 0$. (On pourra pour simplifier chercher un champ qui soit invariant par translation dans la direction verticale, de façon à se ramener à un champ bidimensionnel).

e) Donner un exemple de champ de vitesses u non trivial défini sur \mathbb{R}^3 tel qu'en tout point, la décomposition (V.2.2) soit telle que $\omega = 0$ (champ irrotationnel).

f) (★) Que peut-on dire, au vu de ce qui précède, d'un champ qui dérive d'un potentiel, c'est-à-dire un champ qui s'écrit $u = -\nabla\Phi$, où Φ est une fonction scalaire suffisamment régulière pour que u soit différentiable ?

Exercice V.2.8. (Potentiel d'interaction (••))

On considère la fonction $D(\cdot)$ qui à $q = (q_1, q_2) \in \mathbb{R}^3 \times \mathbb{R}^3$ (attention, q_1 et q_2 désignent ici des points de \mathbb{R}^3) associe la distance entre les points q_1 et q_2 de \mathbb{R}^3 :

$$D(q) = D(q_1, q_2) = \|q_2 - q_1\|.$$

a) Montrer que $D(\cdot)$ est différentiable sur l'ouvert

$$U = \{ q = (q_1, q_2) \in \mathbb{R}^6, q_1 \neq q_2 \},$$

et exprimer son gradient (on pourra exprimer les gradients partiels ∇_{q_1} et ∇_{q_2}) en fonction du vecteur unitaire $e_{12} = (q_2 - q_1) / \|q_2 - q_1\|$.

b) On introduit un potentiel d'interaction sur le système de deux particules localisées en q_1 et q_2 sous la forme $V = V(q) = V(q_1, q_2) = \varphi(D(q_1, q_2))$, où φ est une fonction continûment dérivable de \mathbb{R}_+ dans \mathbb{R} . Montrer que V est différentiable sur U , et écrire son gradient.

c) Préciser les gradients partiels $\nabla_{q_1} V$ et $\nabla_{q_2} V$ si l'on prend pour φ le potentiel d'interaction gravitationnelle défini par $\varphi(D) = -1/D$.

d) On se replace dans le cas général d'un potentiel φ quelconque, et l'on considère maintenant un système de N particules dans \mathbb{R}^3 . On définit un potentiel d'interaction global de la façon suivante

$$V(q) = V(q_1, q_2, \dots, q_N) = \sum_{1 \leq i < j \leq N} \varphi(D(q_i, q_j)).$$

On s'intéresse au système résultant du principe fondamental de la dynamique, sous l'hypothèse de forces dérivant d'un potentiel (on prend des masses unitaires), c'est-à-dire

$$\frac{d^2 q}{dt^2} = -\nabla V(q). \quad (\text{V.2.3})$$

Écrire l'équation qui résulte de ce principe pour chacune des particules, qui s'écrit de façon abstraite

$$\frac{d^2 q_i}{dt^2} = -\nabla_{q_i} V(q).$$

e)(★) On se place dans le cadre des notations de la question précédente. On suppose que l'on connaît une solution $t \in [0, T[\mapsto q(t) \in U$ de l'équation d'évolution (V.2.3). Montrer que l'on a conservation de l'énergie totale, c'est à dire que la quantité

$$E(t) = \sum_{i=1}^N \frac{1}{2} \left\| \frac{dq_i}{dt}(t) \right\|^2 + V(q(t))$$

est constante sur $[0, T[$.

Dans le cas du potentiel gravitationnel $\varphi(D) = -1/D$, peut-on en déduire que les vitesses sont majorées sur $[0, T[$? Même question pour le cas du potentiel coulombien entre charges identiques $\varphi(D) = 1/D$.

Exercice V.2.9. (Densité gaussienne)

a) Calculer

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda(x, y) = \int_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy.$$

(On pourra utiliser les coordonnées polaires sur \mathbb{R}^2 .)

b) Montrer que l'intégrale ci-dessus peut s'exprimer simplement en fonction de $\int_{\mathbb{R}} e^{-x^2} dx$.

c) Pour $\sigma > 0$ fixé, on considère l'application

$$x = (x_1, \dots, x_d) \mapsto f(x) = e^{-\frac{x_1^2 + \dots + x_d^2}{2\sigma^2}}.$$

Calculer le coefficient C assurant que la mesure associée à la fonction f , vue comme une densité, soit de masse totale égale à 1.

V.3 Théorèmes des fonctions implicites et d'inversion locale

Le résultat principal de cette section est le théorème dit *des fonctions implicites*, que l'on peut interpréter comme suit. On considère une équation portant sur $y \in \mathbb{R}^m$, équation qui dépend de *paramètres* x_1, \dots, x_n , et que l'on écrit

$$f(x, y) = 0.$$

Cette équation est à valeurs vectorielles. Pour se placer dans un contexte où l'équation, pour un jeu de paramètres x fixé, peut permettre de déterminer y , on s'intéresse au cas où il y a autant d'équations que d'inconnues, c'est à dire que f est à valeurs dans \mathbb{R}^m . L'inconnue y est donc définie de façon *implicite* par rapport aux paramètres x_1, \dots, x_n . On se place au voisinage d'une solution de cette équation : pour un jeu de paramètres $x = (x_1, \dots, x_n)$ donné, on suppose connue une solution $y = (y_1, \dots, y_m)$ de l'équation. Si l'on fait varier les paramètres de l'équation, on peut s'attendre à ce que, sous certaines conditions, la solution en y varie elle-même de façon régulière. Le théorème ci-dessous donne des conditions suffisantes pour que l'on puisse en effet exprimer y en fonction de x , de façon régulière, au voisinage d'un couple paramètres - solution (x_0, y_0) donné. La condition principale permettant cette explicitation de la dépendance apparaît clairement dans l'exemple-jouet suivant :

$$f : (x, y) \in \mathbb{R}^2 \mapsto ax + by + c.$$

On peut exprimer y fonction de x si et seulement si $b \neq 0$, où b quantifie la manière dont f varie vis-à-vis de y . Dans le cas le plus général ($y \in \mathbb{R}^m$, f à valeurs dans \mathbb{R}^m), cette dépendance sera encodée par la différentielle de f par rapport à y (qui est bien représentée dans la base canonique par une matrice carrée). L'hypothèse principale porte sur le caractère *inversible* de cette différentielle.

Théorème V.3.1. (••) Soit f une fonction définie sur un ouvert W de $\mathbb{R}^n \times \mathbb{R}^m$, à valeurs dans \mathbb{R}^m . On suppose f continûment différentiable sur W , et l'on suppose que la différentielle partielle de f par rapport à y , notée $\partial_y f(x, y)$, est inversible en tout point¹⁰ de W . On considère un point $(x_0, y_0) \in W$ qui annule f :

$$f(x_0, y_0) = 0.$$

On peut alors exprimer y comme fonction de x au voisinage de (x_0, y_0) . Plus précisément : il existe des voisinages ouverts $U \subset \mathbb{R}^n$ et $V \subset \mathbb{R}^m$ de x_0 et y_0 , respectivement, et une fonction Ψ de U dans V , tels que

$$(x, y) \in U \times V, f(x, y) = 0 \iff y = \Psi(x).$$

La fonction Ψ est continûment différentiable sur U , et sa différentielle s'exprime

$$d\Psi(x) = -(\partial_y f(x, y))^{-1} \circ \partial_x f(x, y), \text{ avec } y = \Psi(x).$$

10. Comme précisé dans la remarque V.3.3 ci-après, il suffit de vérifier que la différentielle soit inversible en (x_0, y_0) pour qu'elle le soit dans un voisinage de ce point.

Démonstration. La démarche, de nature constructive, est basée sur un processus itératif construit selon les principes suivants. On considère x proche de x_0 (dans un sens précisé plus loin), et l'on cherche y tel que $f(x, y) = 0$. On suppose que l'on dispose d'une première approximation y_k du y recherché, et on cherche un y_{k+1} qui en soit une meilleure approximation. On a

$$f(x, y_{k+1}) = f(x, y_k + (y_{k+1} - y_k)) \approx f(x, y_k) + \partial_y f(x, y_k) \cdot (y_{k+1} - y_k).$$

On souhaite annuler cette quantité, ce qui suggère de définir y_{k+1} comme

$$y_{k+1} = y_k - (\partial_y f(x, y_k))^{-1} \cdot f(x, y_k).$$

Il s'agit de la méthode dite de *Newton* pour trouver le zéro d'une fonction. Nous allons considérer ici une version modifiée de cette méthode, en remplaçant la différentielle partielle en y par sa valeur au point (x_0, y_0) . Partant de y_0 (en fait, on peut partir d'une valeur initiale différente de y_0 , mais nous le fixons comme point de départ pour simplifier), on construit donc la suite (y_k) par récurrence, selon la formule

$$y_{k+1} = y_k - Q^{-1} \cdot f(x, y_k), \text{ avec } Q = \partial_y f(x_0, y_0).$$

N.B. : On prendra garde au fait que, pour (x, y) donné, $\partial_y f(x, y)$ est une application linéaire de \mathbb{R}^m dans \mathbb{R}^m . Cette application dépend du point $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ où elle est prise, mais sans que la différentielle soit prise par rapport à la variable x . Cette différentielle partielle est définie par le développement limité suivant, où l'on ne perturbe que la variable y : pour $h \in \mathbb{R}^m$,

$$f(x, y + h) = f(x, y) + \partial_y f(x, y) \cdot h + o(h).$$

Il s'agit donc d'un champ d'applications linéaires, auquel on peut associer un champ de matrices carrées $m \times m$ (leurs représentations dans la base canonique de \mathbb{R}^m), qui vit sur un espace de dimension $n \times m$. L'application Q est simplement la valeur particulière de ce champ au point (x_0, y_0) .

Cette récurrence peut s'écrire $y_{k+1} = \Phi_x(y_k)$, où la fonction Φ_x est définie par

$$y \longmapsto \Phi_x(y) = y - Q^{-1} \cdot f(x, y),$$

pour tout y tel que $(x, y) \in W$. Noter que y est point fixe de Φ_x si et seulement si $f(x, y) = 0$. Nous allons montrer que cette fonction admet bien un unique point fixe sur un voisinage de y_0 . Cette fonction est différentiable sur son domaine de définition, de différentielle

$$d\Phi_x(y) = I - Q^{-1} \circ \partial_y f(x, y).$$

En écrivant $I = Q^{-1}Q$ on obtient

$$\|d\Phi_x(y)\| = \|Q^{-1}(\partial_y f(x_0, y_0) - \partial_y f(x, y))\| \leq \|Q^{-1}\| \|\partial_y f(x_0, y_0) - \partial_y f(x, y)\|$$

Fixons $\kappa = 1/2$. La différentielle étant continue, il existe un $r > 0$ tel que, pour tout point $x \in B_f(x_0, r)$ (boule fermée de centre x_0 et de rayon r), tout $y \in B_f(y_0, r)$ (on prend r suffisamment petit pour que $B_f(x_0, r) \times B_f(y_0, r) \subset W$),

$$\|\partial_y f(x_0, y_0) - \partial_y f(x, y)\| \leq \kappa \|Q^{-1}\|^{-1},$$

de telle sorte que

$$\forall x \in B_f(x_0, r), y \in B_f(y_0, r), \|d\Phi_x(y)\| \leq \kappa.$$

On a donc, pour tous y, y' dans $B_f(y_0, r)$,

$$\|\Phi_x(y) - \Phi_x(y')\| \leq \kappa \|y - y'\|$$

d'après le théorème des accroissements finis (théorème V.1.20, page 90), avec $\kappa = 1/2$. L'application Φ_x est donc contractante sur $B_f(y_0, r)$. Montrons qu'elle laisse stable une boule autour de y_0 . Comme l'application

$$x \longmapsto \Phi_x(y_0) = y_0 - Q^{-1} \cdot f(x, y_0)$$

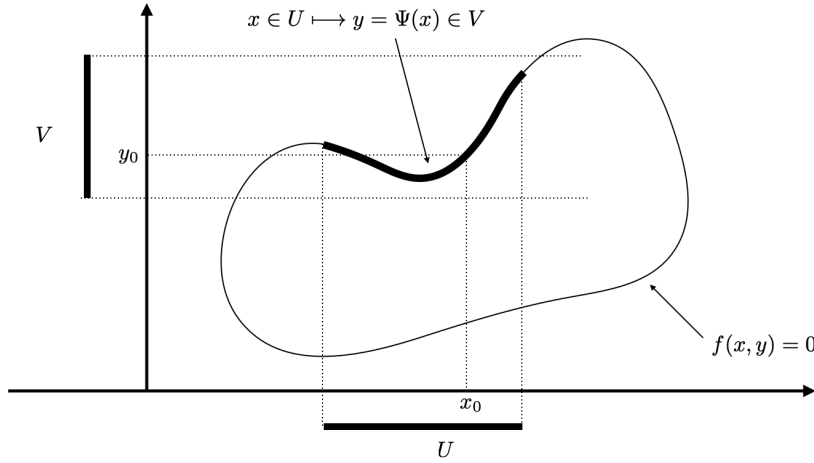


FIGURE V.3.1 – Théorème des fonctions implicites

est continue en x_0 , il existe un $r' < r$ tel que, pour tout $x \in B_f(x_0, r')$, on ait

$$\|\Phi_x(y_0) - \Phi_{x_0}(y_0)\| \leq (1 - \kappa)r,$$

avec $\Phi_{x_0}(y_0) = y_0$ car $f(x_0, y_0) = 0$. On a alors, pour tout $x \in B_f(x_0, r')$, tout $y \in B_f(y_0, r)$,

$$\|\Phi_x(y) - y_0\| = \|\Phi_x(y) - \Phi_{x_0}(y_0)\| \leq \underbrace{\|\Phi_x(y) - \Phi_x(y_0)\|}_{\leq \kappa\|y - y_0\|} + \underbrace{\|\Phi_x(y_0) - y_0\|}_{\leq (1 - \kappa)r} \leq \kappa r + (1 - \kappa)r = r.$$

Pour tout $x \in B_f(x_0, r')$, l'application Φ_x est donc bien définie de $B_f(y_0, r)$ dans lui-même, et cet ensemble est complet comme fermé dans le complet \mathbb{R}^m . Elle par ailleurs contractante comme montré précédemment. D'après le théorème I.8.2, elle admet donc un unique point fixe sur $B_f(y_0, r)$, c'est-à-dire qu'il existe un unique $y \in B_f(y_0, r)$ tel que $f(x, y) = 0$. On note Ψ l'application qui à x associe cette unique solution en y de $f(x, y) = 0$.

Montrons maintenant la continuité de Ψ , et précisons le choix des voisinages U et V . Soient x_1 et x_2 deux points de $B_f(x_0, r')$, et $y_1 = \Psi(x_1)$, $y_2 = \Psi(x_2)$. On a

$$\|y_2 - y_1\| = \|\Phi_{x_2}(y_2) - \Phi_{x_1}(y_1)\| \leq \|\Phi_{x_2}(y_2) - \Phi_{x_2}(y_1)\| + \|\Phi_{x_2}(y_1) - \Phi_{x_1}(y_1)\|.$$

Comme Φ_{x_2} est κ -contractante sur $B_f(y_0, r)$, on a $\|\Phi_{x_2}(y_2) - \Phi_{x_2}(y_1)\| \leq \kappa\|y_2 - y_1\|$, d'où

$$\begin{aligned} \|y_2 - y_1\| &\leq \frac{1}{1 - \kappa} \|\Phi_{x_2}(y_1) - \Phi_{x_1}(y_1)\| = \frac{1}{1 - \kappa} \|Q^{-1} \cdot (f(x_2, y_1) - f(x_1, y_1))\| \\ &\leq \frac{1}{1 - \kappa} \|Q^{-1}\| \max_{B_f(x_0, r') \times B_f(y_0, r)} \|\partial_x f\| \|x_2 - x_1\| \end{aligned}$$

d'après le théorème des accroissements finis V.1.20 (f étant continûment différentiable sur le compact $B_f(x_0, r') \times B_f(y_0, r)$, sa différentielle partielle par rapport à x est bornée). vers 0 quand x_2 tend vers x_1 . L'application Ψ est donc continue sur $B_f(x_0, r')$ à valeurs dans $B_f(y_0, r)$, et même lipschitzienne : il existe $C > 0$ tel que

$$\|\Psi(x_2) - \Psi(x_1)\| \leq C \|x_2 - x_1\|.$$

Soit V voisinage ouvert de y_0 inclus dans $B_f(y_0, r)$. Comme Ψ est continue, il existe un voisinage ouvert de x_0 , $U \subset B_f(x_0, r')$, tel que $\Psi(U) \subset V$.

Il reste à montrer que Ψ est différentiable sur U . Soit $x \in U$, $y = \Psi(x) \in V$. On considère une variation h de x telle que $x + h \in U$. Il existe un unique g tel que $y + g \in V$ vérifie

$$f(x + h, y + g) = 0.$$

D'après ce qui précède il existe $C > 0$ tel que $\|g\| \leq C \|h\|$. La différentiabilité de f en (x, y) s'exprime

$$\underbrace{f(x+h, y+g)}_{=0} = \underbrace{f(x, y)}_{=0} + \partial_x f(x, y) \cdot h + \partial_y f(x, y) \cdot g + o(h, g).$$

On a donc

$$g = - \left((\partial_y f(x, y))^{-1} \circ \partial_x f(x, y) \right) \cdot h + o(h),$$

(le $o(h, g)$ s'est bien transformé en $o(h)$ du fait que la norme de h domine celle de g , comme indiqué précédemment). Ce g est, par construction, $\Psi(x+h) - \Psi(x)$, on a donc

$$\Psi(x+h) = \Psi(x) - \left((\partial_y f(x, y))^{-1} \circ \partial_x f(x, y) \right) \cdot h + o(h).$$

ce qui exprime que l'application $x \mapsto \Psi(x)$ est différentiable sur U , de différentielle

$$d\Psi(x) = (\partial_y f(x, \Psi(x)))^{-1} \circ \partial_x f(x, \Psi(x)).$$

Comme f est continûment différentiable, et que Ψ est continue, $x \mapsto d\Psi(x)$ est continue. \square

Remarque V.3.2. On notera que, par construction, Ψ est bien définie sur tout U mais, comme illustré par la figure V.3.1, elle n'est pas nécessairement surjective (cette remarque sera importante pour la démonstration du théorème des fonctions implicites, dans lequel il s'agira de construire deux ouverts en bijection). Par ailleurs, pour $x \in U$, il peut exister plusieurs y tels que $f(x, y) = 0$, mais un seul qui soit dans V .

Remarque V.3.3. Pour vérifier l'applicabilité du théorème précédent en un point (x_0, y_0) qui annule f , et au voisinage duquel f est définie, il suffit de vérifier que la différentielle de f par rapport à y est inversible en (x_0, y_0) . En effet, si c'est le cas, l'application $(x, y) \mapsto \partial_y f(x, y)$ étant continue, et le déterminant étant une fonction continue, la différentielle reste inversible sur un ouvert de (x_0, y_0) , qui peut jouer le rôle du W dans les hypothèses du théorème précédent. On dira que le théorème des fonctions implicites s'applique en (x_0, y_0) , ou *au voisinage de* (x_0, y_0) .

Remarque V.3.4. Ce théorème, qui peut sembler assez abstrait et technique, peut être invoqué d'une manière *négative* pour qualifier la pertinence d'un modèle. Replaçons-nous dans le cadre de l'introduction, en interprétant $f(x, y)$ comme un *modèle* portant sur y , sous la forme d'un système d'équations dépendant de paramètres x_1, \dots, x_n . Le modèle a vocation à, pour un jeu de paramètres (qui peuvent être des températures, des pressions, des flux d'information, des prix, ...), déterminer la collection des inconnues y_1, \dots, y_m . Dans le cadre d'une utilisation de ce modèle dans la vie réelle, les paramètres ne sont en général connus qu'approximativement (erreurs de mesure, variabilité en temps de paramètres supposés statiques, ...). Si la solution y ne dépend *pas* de façon régulière des paramètres, cela signifie qu'une erreur petite sur les paramètres peut induire une variation très importante de la solution. On dira que le problème n'est pas *stable*¹¹. Du fait de la non-différentiabilité de la correspondance paramètres \mapsto solution (même si le problème est bien posé au sens où la solution est définie de façon unique), il n'existera pas de constante c telle qu'une erreur relative ε sur les paramètres induise une erreur contrôlée par $c\varepsilon$. Un tel modèle est essentiellement *inutilisable* en situation réelle, ou tout du moins très délicat à exploiter.

Remarque V.3.5. (Sensibilité vis à vis des paramètres)

Dans la continuité de la remarque précédente, mais de façon plus positive, lorsque l'on est bien dans le cadre du théorème des fonctions implicites, la différentielle de Ψ précise la dépendance de la solution vis-à-vis des paramètres. On écrira en général simplement $\Psi(x) = y(x)$, de telle sorte que la matrice jacobienne de Ψ contient les dérivées partielles $\partial y_i / \partial x_j$ (où y est maintenant considéré comme fonction de x), c'est-à-dire l'expression de la dépendance de la i -ième composante de y vis-à-vis du paramètre x_j (on parlera de *sensibilité*). Les paramètres les plus significatifs pour une composante y_i correspondent aux fortes valeurs de la dérivée, il sera important de bien maîtriser la valeur, alors que les paramètres pour lesquels $\partial y_i / \partial x_j$

11. On parle parfois de stabilité au sens de Hadamard, même si cette appellation fait plutôt référence à une dépendance *continue* de la solution par rapport aux données.

est petit pour tous les i peuvent être a priori estimés avec une précision médiocre, sans que cela n'influe de façon préjudiciable sur la solution.

Remarque V.3.6. (Identification de paramètres)

Il est courant de s'intéresser au *problème inverse*, qui peut se formuler comme suit. On fait confiance au modèle $f(x, y) = 0$, on dispose de mesures pour la solution y , et l'on cherche à *estimer les paramètres* correspondant à la solution mesurée. On est donc amené à considérer le problème dans l'autre sens, c'est-à-dire que l'on cherche à estimer x à partir de la connaissance de y . On ne peut espérer retrouver exactement les paramètres que si leur nombre est égal à celui des inconnues $n = m$. On notera que, pour ce nouveau problème, les paramètres les plus difficiles à identifier précisément sont ceux qui ont peu d'influence sur la solution, qui étaient considérés pour le problème direct comme peu significatifs, dont la connaissance précise n'était pas nécessaire. C'est précisément leur peu d'influence sur la solution qui rend difficile leur estimation à partir de la connaissance de cette solution¹².

Définition V.3.7. Soit φ une application d'un ouvert $U \subset \mathbb{R}^n$ dans un ouvert $V = \varphi(U)$ dans \mathbb{R}^n . On dit que φ est un C^1 – difféomorphisme de U vers V si φ est bijective, et si φ et sa réciproque φ^{-1} sont continûment différentiables.

Proposition V.3.8. On se place dans les hypothèses de la définition précédente. La différentielle de φ est inversible en tout point de U , et son inverse est la différentielle de l'application réciproque φ^{-1} : pour tout $x \in U$, $y = \varphi(x) \in V$,

$$d\varphi^{-1}(y) = (d\varphi(x))^{-1}.$$

Démonstration. On a, pour tout $y \in V$,

$$\varphi \circ \varphi^{-1}(y) = y.$$

La règle de différentiation en chaîne implique donc (avec $x = \varphi^{-1}(y)$)

$$d\varphi(x) \circ d\varphi^{-1}(y) = \text{Id},$$

qui conclut la preuve. □

Théorème V.3.9. (Inversion locale)

Soit φ une application continûment différentiable d'un ouvert $W \subset \mathbb{R}^n$ dans \mathbb{R}^n . On suppose que $d\varphi(x)$ est inversible pour tout $x \in W$. Alors φ est un C^1 – difféomorphisme local : pour tout $x_0 \in W$, il existe un voisinage ouvert $U \subset W$ de x_0 et un voisinage ouvert V de $y_0 = \varphi(x_0)$ tel que $\varphi|_U$ soit un C^1 – difféomorphisme de U vers V .

Démonstration. On considère l'application (noter que l'on écrit (y, x) du fait qu'il va s'agir, contrairement à l'usage, d'exprimer x en fonction de y) :

$$g : (y, x) \in \mathbb{R}^n \times W \mapsto g(y, x) = \varphi(x) - y.$$

Cette application est différentiable sur $\mathbb{R}^n \times W$, de différentielle partielle par rapport à x

$$\partial_x g(y, x) = d\varphi(x).$$

Cette différentielle est inversible sur W par hypothèse. Soit $x_0 \in W$, et $y_0 = \varphi(x_0)$, d'où $g(y_0, x_0) = 0$. D'après le théorème des fonctions implicites, il existe un voisinage V de y_0 , un voisinage $\tilde{U} \subset W$ de x_0 , et Ψ une application continûment différentiable de V dans \tilde{U} , tels que

$$(y, x) \in V \times \tilde{U}, \quad g(y, x) = 0, \quad \text{i.e.} \quad y = \varphi(x) \iff x = \Psi(y).$$

12. Nous nous en tenons dans cette remarque à une vision un peu simpliste des choses, comme s'il était possible de séparer à la fois les paramètres et les composantes de la solution (ça n'est possible que si la différentielle est diagonale). En tout généralité, les études de sensibilité évoquées dans ces remarques passent par une étude plus complète de la matrice dans sa globalité, qui passe en particulier par une analyse spectrale.

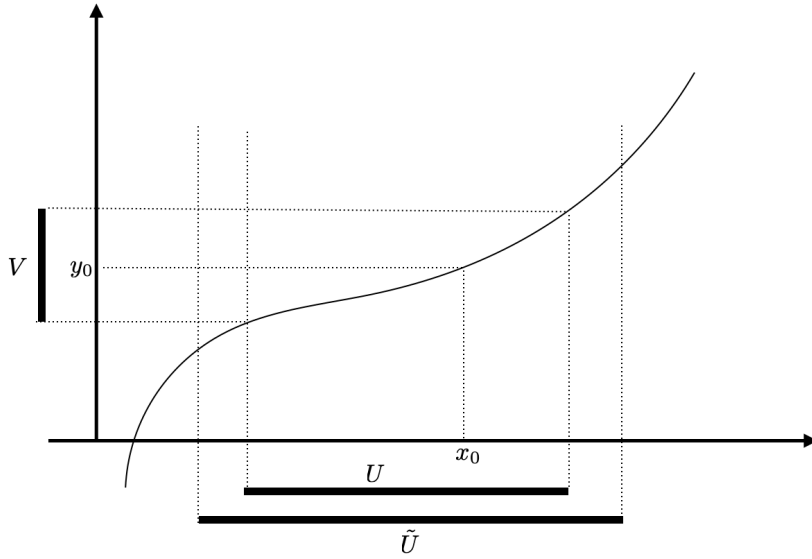


FIGURE V.3.2 – Théorème d'inversion locale

L'application Ψ est donc la réciproque de φ . Il reste à préciser les voisinages ouverts de x_0 et y_0 qui sont en bijection. Il faut prendre garde à une difficulté (annoncée dans la remarque V.3.2) : Ψ , qui est bien définie sur tout V , (cet ouvert V est noté U dans le théorème des fonctions implicites, du fait du renversement des rôles de x et y que nous avons effectué ici), n'est pas nécessairement *surjective* de V dans \tilde{U} . Pour garantir que les deux ouverts soient en bijection, on réduit l'ouvert \tilde{U} en introduisant

$$U = \tilde{U} \cap \Psi(V).$$

Comme, pour tout $y \in V$, l'équation $y = \varphi(x)$ n'a qu'une solution en $x \in \tilde{U}$, cet ensemble s'écrit aussi $U = \tilde{U} \cap \varphi^{-1}(V)$. Il s'agit donc bien d'un ouvert par continuité de V . \square

V.4 Problème adjoint

On s'intéresse ici à une méthode d'estimation du gradient d'une fonctionnelle $F(u)$ à valeurs réelles, définie comme $g(y_u)$, où y_u est solution d'un problème de type

$$f(u, y) = 0. \quad (\star)$$

considéré comme un problème en y , paramétré par u . On assimilera dans la suite la différentielle d'une application à la matrice jacobienne qui la représente dans les bases canoniques sur les espaces d'arrivée et de départ.

On considère une application f de $\mathbb{R}^n \times \mathbb{R}^m$ dans \mathbb{R}^m , continument différentiable sur un ouvert W de $\mathbb{R}^n \times \mathbb{R}^m$. On notera $u \in \mathbb{R}^n$ la variable dite *de contrôle*, et $y \in \mathbb{R}^m$ la variable d'état.

On considère $(u_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ solution de $f(u_0, y_0) = 0$, et l'on suppose que $\partial_y f(u, y)$ est inversible dans un voisinage¹³ de (u_0, y_0) . D'après le théorème des fonctions implicites V.3.1, il existe un ouvert U contenant u_0 et un ouvert V contenant y_0 , et une application Ψ de U dans V telle que

$$\forall (u, y) \in U \times V, \quad f(u, y) = 0 \iff y = \Psi(u),$$

13. Il suffit en fait de supposer que $\partial_y f(u_0, y_0)$ est inversible. La matrice qui la représente dans la base canonique de \mathbb{R}^m est de déterminant non nul. Comme le déterminant est fonction continue des coefficients, le déterminant reste non nul, et donc $\partial_y f(u, y)$ est inversible, dans un voisinage W de (u_0, y_0) .

avec Ψ continument différentiable sur U . La différentielle de Ψ en $u \in U$ s'exprime

$$d\Psi(u) = -(\partial_y f(u, y_u))^{-1} \circ \partial_u f(u, y_u), \quad (\text{V.4.1})$$

avec $y_u = \Psi(u)$. Par la suite, pour désigner le y associé à u comme solution de l'équation (\star) , on utilisera indifféremment $\Psi(u)$ ou y_u .

On considère maintenant une application g de \mathbb{R}^m dans \mathbb{R} , et l'on définit la fonction F de $U \subset \mathbb{R}^n$ dans \mathbb{R} par

$$F(u) = g(y_u) = g \circ \Psi(u),$$

où y_u est la solution de (\star) associée à u .

Approche par inversion de matrice

La différentielle $dF(u)$ s'exprime .

$$dF(u) = dg(y_u) \circ d\Psi(u) = -dg(y_u) \circ (\partial_y f(u, y_u))^{-1} \circ \partial_u f(u, y_u). \quad (\text{V.4.2})$$

Exercice V.4.1. On se propose d'appliquer ce qui précède à la situation suivante : on suppose $n = 1$, $m \geq 1$, et l'on note $A(u)$ une matrice de $\mathcal{M}_m(\mathbb{R})$ dont les coefficients dépendent de la variable scalaire u :

$$A(u) = (a_{ij}(u))_{1 \leq i, j \leq m}.$$

On suppose que les $u \mapsto a_{ij}(u)$ sont des fonctions continument différentiables sur \mathbb{R} . Le problème définissant y fonction de u s'écrit sous la forme d'un système matriciel en y

$$A(u) \cdot y = b,$$

où b est un vecteur donné de \mathbb{R}^m . On suppose que $A(u_0)$ est inversible pour un certain $u_0 \in \mathbb{R}$. On définit g comme

$$y \in \mathbb{R}^m \mapsto g(y) = \frac{1}{2} \|y - \bar{y}\|^2,$$

où $\bar{y} \in \mathbb{R}^m$ est donné.

a) On pose $f(u, y) = A(u) \cdot y - b$.

Préciser les expressions de $\partial_y f$ et $\partial_u f$ en fonction de la matrice A et de sa dérivée $A' = (a'_{ij})$.

N.B. : On pourra effectuer pour cela les développements limités de $f(u + \delta u, y)$ et $f(u, y + \delta y)$

b) Montrer que ce problème rentre dans le cadre général décrit au début de cette section, en déduire que l'on peut exprimer y en fonction de u sur un voisinage U de u_0 , sous la forme $y = \Psi(u)$, et donner l'expression de la différentielle de Ψ .

N.B. : cette différentielle étant une application de \mathbb{R} dans \mathbb{R}^m , on pourra la représenter par une matrice colonne.

c) Préciser la différentielle de l'application $y \in \mathbb{R}^m \mapsto g(y) = \frac{1}{2} \|y - \bar{y}\|^2$, en expliquant pourquoi on peut l'identifier à une matrice ligne. Quel est le lien entre cette matrice ligne et le gradient de g (pour le produit scalaire canonique sur \mathbb{R}^m) ?

Expliquer pourquoi la différentielle de $u \mapsto F(u)$ peut être assimilée à un réel, et préciser l'expression de cette différentielle en u_0 .

Approche par différences finies

Si l'on souhaite estimer la différentielle de F en u_0 , sans avoir à inverser la matrice $\partial_y f(u, y_u)$, on peut approcher chacune des dérivées partielles par

$$\partial_{u_j} F(u) \approx \frac{F(u + h e_j) - F(u)}{h},$$

pour un certain $h > 0$ petit, où les e_i sont les vecteurs de la base canonique de \mathbb{R}^n . Chaque estimation de F en un point nécessite une résolution d'un problème de type (\star) , et l'on doit estimer F en u_0 et en les $u + h e_j$, ce qui nécessite $n + 1$ résolutions.

Approche par méthode de l'état adjoint

Nous présentons maintenant une méthode d'estimation de $dF(u)$ (appelée méthode de l'état adjoint) qui ne nécessite ni de calculer l'inverse d'une matrice (approche de la question 2a basée sur l'expression de la différentielle), ni de résoudre un grand nombre de problèmes de type (\star) (approche par différences finies décrite précédemment).

On définit l'application L (appelée *Lagrangien*) de la façon suivante

$$L : (y, u, p) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \longmapsto L(y, u, p) = g(y) + \langle f(u, y) | p \rangle.$$

Noter que, dans l'expression ci-dessus, y est une variable "libre" (il ne s'agit pas de y_u).

Par définition de y_u , quel que soit p , on $F(u) = L(y_u, u, p)$, avec $y_u = \Psi(u)$. Si l'on différencie cette identité par rapport à u , à p fixé quelconque, on obtient¹⁴

$$dF(u) = \partial_y L(y_u, u, p) \circ d\Psi(u) + \partial_u L(y_u, u, p) \quad \forall p \in \mathbb{R}^m.$$

On se place en un point u_0 de l'espace des paramètres, et l'on note, pour alléger, $y_0 = y_{u_0} = \Psi(u_0)$ l'état associé. On appelle *problème adjoint* le problème consistant à trouver p tel que

$$\partial_y L(y_0, u_0, p) = 0, \quad \text{avec } L(y_0, u_0, p) = g(y_0) + \langle f(u_0, y_0) | p \rangle.$$

Précisons la forme de ce problème adjoint. Le premier terme de $\partial_y L(y_0, u_0, p)$ s'écrit simplement $dg(y_0)$. Pour le second terme, on écrit

$$\langle f(u, y + \delta y) | p \rangle = \langle f(u, y) | p \rangle + \langle \partial_y f(u, y) \cdot \delta y | p \rangle = \langle \partial_y f(u, y)^T \cdot p | \delta y \rangle,$$

où l'on a assimilé l'application linéaire $\partial_y f(u, y)$ à sa matrice dans la base canonique, $\partial_y f(u, y)^T$ représente la matrice transposée, et le ' \cdot ' le produit matrice-vecteur. On assimilera la différentielle de $\langle f(u, y) | p \rangle$ à son gradient pour le produit scalaire canonique, on écrira donc $\partial_y f(u, y)^T \cdot p$ la différentielle en y de $\langle f(u, y) | p \rangle$.

On a donc

$$\partial_y L(y_0, u_0, p) = dg(y_0) + \partial_y f(u_0, y_0)^T \cdot p.$$

Le problème adjoint s'écrit ainsi

$$\partial_y f(u_0, y_0)^T \cdot p = -dg(y_0).$$

Il s'agit d'un système linéaire, qui fait intervenir la matrice carrée de $\partial_y f(u_0, y_{u_0})^T \in \mathcal{M}_m(\mathbb{R})$. Comme c'est la matrice transposée de $\partial_y f(u_0, y_{u_0})$, qui est inversible, elle est elle-même inversible, le problème admet donc une solution unique.

Nous pouvons maintenant préciser la stratégie qui permet d'exprimer $dF(u_0)$ sans avoir à calculer $d\Psi(u_0)$. Nous avons montré

$$dF(u_0) = \partial_y L(y_0, u_0, p) \circ d\Psi(u_0) + \partial_u L(y_0, u_0, p) \quad \forall p \in \mathbb{R}^m.$$

Cette expression est en particulier vraie pour la solution p_0 du problème adjoint :

$$\partial_y f(u_0, y_0)^T \cdot p_0 = -dg(y_0). \tag{V.4.3}$$

Pour ce p_0 particulier, on a $\partial_y L(y_{u_0}, u_0, p_0) = 0$, ce qui annule le premier terme de l'expression de $dF(u)$ ci-dessus, d'où

$$dF(u_0) = \underbrace{\partial_y L(y_0, u_0, p_0)}_{=0} \circ d\Psi(u_0) + \partial_u L(y_0, u_0, p_0) = \partial_u L(y_0, u_0, p_0),$$

ce qui permet bien de s'affranchir du calcul de $d\Psi(u_0)$.

14. On prendra garde au fait que $\partial_u L(y_u, u, p)$ désigne bien la différentielle de L par rapport à la seconde variable (u) uniquement, prise au point (y_u, u, p) .

Remarque V.4.1. L'intérêt apporté par cette démarche par rapport à l'utilisation brute de la formule (V.4.1), qui implique l'inverse d'une matrice $m \times m$ n'est pas si évident, du fait que le problème adjoint repose lui-même sur la résolution d'un système $m \times m$ (équation V.4.3). De fait, si le nombre de paramètres est petit (comme dans l'exemple de l'exercice V.4.2 ci-après), l'approche par état adjoint n'apporte rien en termes d'efficacité. Mais si le nombre n de paramètre est important, le calcul de $(\partial_y f(u, y_u))^{-1} \circ \partial_u f(u, y_u)$ dans l'approche brute nécessite¹⁵ la résolution de n systèmes $m \times m$, alors que, avec l'approche par état adjoint, on n'a qu'un seul système à résoudre (équation V.4.3).

Exercice V.4.2. Appliquer la démarche décrite précédemment à la situation particulière de l'exercice V.4.1 ci-dessus, où, nous le rappelons, u est une variable *scalaire*. On précisera en particulier le problème adjoint, dont on justifiera le caractère bien posé, et on donnera l'expression de $dF(u)$ qui fait intervenir la solution à ce problème adjoint.

Exercice V.4.3. (Méthode de “backpropagation” pour les réseaux de neurones)

On s'intéresse ici à un réseau de neurones, défini par la donnée de p applications (non linéaires a priori), qui sont les *couches* du réseau :

$$\begin{aligned} y_0 \in \mathbb{R}^{m_0} &\longmapsto f_1(y_0, u_1) \in \mathbb{R}^{n_1}, \quad y_1 \in \mathbb{R}^{m_1} \longmapsto f_2(y_1, u_2) \in \mathbb{R}^{m_2}, \\ \dots, \quad y_{p-1} \in \mathbb{R}^{m_{p-1}} &\longmapsto f_p(y_{p-1}, u_p) \in \mathbb{R}^{m_p}, \quad y_i \in \mathbb{R}^{m_i}, \quad u_i \in \mathbb{R}^{n_i} \end{aligned}$$

Pour une collection $u = (u_1, \dots, u_p) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$ de paramètres (les “poids” des différentes couches du réseau de neurones), pour tout $x \in \mathbb{R}^{m_0}$ donné, on applique successivement les différentes couches : $y_0 = x$, $y_1 = f_1(y_0, u_1)$, \dots , $y_p = f_p(y_{p-1}, u_p)$. On note $\Phi(u, x)$ le $y_p \in \mathbb{R}^{m_p}$ obtenu en fin de chaîne, qui est une fonction de x qui dépend de l'ensemble des paramètres rassemblés en un vecteur $u = (u_1, \dots, u_p)$ dans $\mathbb{R}^{n_1 + \dots + n_p}$. On se donne un certain nombre de données d'entrée $\bar{x}_1, \dots, \bar{x}_K$ dans \mathbb{R}^{m_0} , associées à des données de sorties labellisées $\bar{y}_1, \dots, \bar{y}_K$ dans \mathbb{R}^{m_p} , et l'on introduit

$$u \longmapsto F(u) = \frac{1}{2} \sum_{k=1}^K |\Phi(u, \bar{x}_k) - \bar{y}_k|^2.$$

La phase d'apprentissage d'un tel réseau consiste à trouver des poids u qui minimisent la fonction $F(u)$ (appelée fonction *loss* dans ce contexte) définie ci-dessus. Dans cette optique, il est précieux de pouvoir estimer efficacement le gradient de F . En vous inspirant de la démarche décrite précédemment, proposer une stratégie permettant d'estimer ce gradient en un point u donné. On pourra faire les hypothèses de régularité jugées nécessaires pour appliquer ce qui précède.

V.5 Exercices

Exercice V.5.1. Soit f une fonction continûment différentiable sur $\mathbb{R}^n \times \mathbb{R}^m$, à valeurs dans \mathbb{R}^m .

Montrer que l'ensemble $F = \{(x, y), f(x, y) = 0\}$ est un fermé de $\mathbb{R}^n \times \mathbb{R}^m$, et que l'ensemble des (x_0, y_0) au voisinage desquels on peut appliquer le théorème des fonctions implicites est un ouvert du fermé F (pour la topologie induite, dont les ouverts sont les intersections d'ouverts de $\mathbb{R}^n \times \mathbb{R}^m$ avec F).

Exercice V.5.2. Identifier dans les cas suivants l'ensemble F des solutions de $f(x, y) = 0$, ainsi que l'ensemble des points (x, y) au voisinage desquels on peut appliquer le théorème des fonctions implicites. Préciser, pour les points en lesquels les hypothèses ne sont pas vérifiées, si intervertir les rôles de x et y permet de les vérifier.

15. Dans la résolution pratique de systèmes linéaires, on ne calcule jamais l'inverse d'une matrice dès que la dimension est importante. Il existe en revanche des méthodes de résolution très efficace pour résoudre des problèmes de type $Ax = b$, pour un second membre b donné. C'est une telle méthode que l'on va utiliser pour résoudre le problème adjoint V.4.3. Pour estimer la matrice $(\partial_y f(u, y_u))^{-1} \circ \partial_u f(u, y_u)$, on évitera l'approche consistant à inverser la matrice carrée et à faire un produit matrice vecteur, on résoudra plutôt n systèmes linéaires impliquant cette matrice, pour des seconds membres qui correspondent aux n colonnes de $\partial_u f(u, y_u)$.

- a) $f(x, y) = x^2 + y^2 - r^2$.
 b) $f(x, y) = y - x^2$.
 c) $f(x, y) = (y - x^3)y$.

Exercice V.5.3. (Coordonnées polaires)

a) Montrer que l'application

$$T : (r, \theta) \in U =]0, +\infty[\times]-\pi, \pi[\mapsto (r \cos \theta, r \sin \theta)$$

est un C^1 difféomorphisme entre U et $V = \mathbb{R}^2 \setminus (\{0\} \times \mathbb{R}_-)$.

b) On considère une fonction f de V dans \mathbb{R} , et l'on note g sa version polaire, i.e. $g(r, \theta) = f(r \cos \theta, r \sin \theta)$. Préciser le lien entre les différentielles de f et de g .

Exercice V.5.4. On note $U = \mathbb{R}^2 \setminus \{(0, 0)\}$. Montrer que l'application

$$\varphi : (x, y) \in U \mapsto (x^2 - y^2, 2xy) \in \mathbb{R}^2$$

est un difféomorphisme local au voisinage de tout point de son domaine, mais pas global.

Exercice V.5.5. (Dépendance d'une racine simple d'un polynôme réel vis-à-vis des coefficients)

A toute collection de coefficients $c = (c_0, c_1, \dots, c_N) \in \mathbb{R}^{N+1}$ on associe le polynôme

$$P_c(X) = c_0 + c_1 X + \dots + c_N X^N.$$

On se donne \tilde{c} et \tilde{z} tels que $\tilde{z} \in \mathbb{R}$ est racine *simple* du polynôme $P_{\tilde{c}}$. Montrer qu'il existe une fonction différentiable Ψ des coefficients, définie dans un voisinage U de \tilde{c} , telle que $\tilde{z} = \Psi(\tilde{c})$ et telle que, pour tout $c \in U$, $z = \Psi(c)$ est racine du polynôme P_c .

Exprimer la différentielle de Ψ .

Exercice V.5.6. (Dépendance d'une racine simple d'un polynôme complexe vis-à-vis des coefficients (version complexe de l'exercice V.5.5))

À toute collection de coefficients $c = (c_0, \dots, c_n) \in \mathbb{C}^n$ on associe le polynôme

$$P_c(X) = c_0 + c_1 X + \dots + c_N X^N.$$

On se donne \tilde{c} et \tilde{z} tels que $\tilde{z} \in \mathbb{C}$ est racine *simple* du polynôme $P_{\tilde{c}}$. Montrer qu'il existe une fonction différentiable Ψ des coefficients, définie dans un voisinage U de \tilde{c} , telle que $\tilde{z} = \Psi(\tilde{c})$ et telle que, pour tout $c \in U$, $z = \Psi(c)$ est racine du polynôme P_c .

Exercice V.5.7. On se propose d'étudier un modèle simplifié de bilan radiatif de la terre, basé sur l'écriture d'un équilibre entre l'énergie solaire reçue par la terre et l'énergie ré-émise par rayonnement, supposé suivre la loi de Stefan-Boltzman. On note F le flux de rayonnement solaire reçu en moyenne par unité de surface sur terre, $F_0 \approx 341 \text{ W m}^{-2}$. On considère qu'une fraction $A \in [0, 1]$ de cette énergie est immédiatement réfléchi, où A_0 , appelé *albedo*, est autour de 0.3. L'énergie émise en moyenne par unité de surface par la terre s'écrit σT^4 , où T est la température moyenne (exprimée en Kelvin), et $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. On considère qu'une fraction de cette énergie n'est pas rayonnée vers l'espace, du fait de l'effet de serre. On note $S \in [0, 1]$ la fraction d'énergie qui n'est pas évacuée vers l'espace. Ce paramètre sans dimension est estimé à $S = 0.4$. En supposant que l'on est à l'équilibre, on écrit le bilan entre les énergies reçues et émises :

$$\sigma T^4 (1 - S) = (1 - A) F.$$

a) Estimer la température moyenne T_0 à la surface de la terre associée aux valeurs de référence S_0 , A_0 et F_0 selon ce modèle, et estimer la valeur qu'aurait cette température s'il n'y avait pas d'effet de serre (en supposant que le modèle reste valide¹⁶).

¹⁶ Vue la baisse de température importante induite par cette suppression virtuelle de l'effet de serre, une grande part de l'eau liquide (peu réfléchissante) à la surface du globe se transformerait en glace (fort pouvoir réfléchissant), ce qui entraînerait une augmentation significative de l'albedo A , qui réduirait encore la température d'équilibre.

b) Montrer¹⁷ que, au voisinage du point d'équilibre considéré, on peut exprimer la température comme une fonction continûment différentiable des paramètres S , A , et F .

Exprimer la différentielle de cette fonction, et en déduire le coefficient de proportionnalité entre une variation de S autour de la valeur S_0 et la variation en degrés de la température. Quelle variation de S induit une augmentation de la température de 2°C ?

Si l'on mesure une petite variation de température de δT autour de T_0 , que peut-on dire (toujours dans l'hypothèse où l'on accorde une foi absolue au modèle) des variations δS , δA , et δF qui ont pu induire cette variation de température ?

c) On estime que le CO_2 est responsable de 60 % de l'effet de serre dû aux Gaz à Effets de Serre (GES), eux-mêmes responsables de 30 % de l'effet de serre global. Si l'on admet que l'effet de serre dû au CO_2 est proportionnel à sa concentration dans l'atmosphère, estimer l'augmentation du taux de CO_2 qui conduirait, selon ce modèle à une augmentation de la température de 2°C .

d) On considère que l'albedo dépend lui même de la température : une augmentation de la température est susceptible d'induire une fonte des glaces, qui diminue la part de surface fortement réfléchissante, d'où une diminution de l'albedo. On écrit donc, pour encoder ce phénomène,

$$A = A_0 - \beta(T - T_0),$$

avec $\beta > 0$ (exprimé en K^{-1}). On considérera par ailleurs le terme F de flux radiatif fixé à sa valeur de référence F_0 . Faire l'étude de ce nouveau modèle au voisinage du point d'équilibre de référence (S_0, T_0) .

e) (★) L'effet de serre, qui dépend par exemple de la masse nuageuse présente en moyenne dans l'atmosphère, dépend lui-même de la température. Explorer la manière dont cette dépendance est susceptible d'affecter les considérations précédentes (on pourra écrire S comme la somme d'un terme dépendant du CO_2 , et d'autres termes susceptibles de dépendre directement de la température).

V.6 Dérivées d'ordre supérieur

Cette section porte sur les dérivées d'ordre supérieur. Nous nous focalisons au départ sur la différentielle seconde d'une fonction scalaire, et sur la notion de *matrice hessienne* qui permet de la représenter dans une base orthonormée, puis nous présentons un cadre plus abstrait permettant de généraliser ces notions à des applications à valeurs dans un espace multidimensionnel, et de définir une notion de dérivation à un ordre arbitraire.

V.6.1 Dérivées partielles d'ordre supérieur pour les fonctions scalaires

Définition V.6.1. (Dérivées partielles d'ordre 2)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} . On suppose que f admet des dérivées partielles $\partial f / \partial x_i$ continues sur U (f est donc continûment différentiable d'après la proposition V.1.8). Si chacune de ces dérivées partielles est dérivable en x par rapport à chacune des variables, on appelle dérivées partielles d'ordre 2 les quantités correspondantes, notées

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right).$$

Définition V.6.2. (Matrice hessienne)

Dans le cadre de la définition précédente, on appelle *matrice hessienne* en x , et l'on note $H_f(x)$ (ou plus simplement $H(x)$ s'il n'y a pas d'ambiguïté) la matrice carrée dont les éléments sont les dérivées partielles d'ordre 2

$$H(x) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(x) \right).$$

¹⁷. Même si cela n'est pas à strictement parler nécessaire ici, on s'efforcera de *jouer le jeu* en utilisant le théorème des fonctions implicites.

La proposition qui suit, capitale, établit que, si les dérivées secondes sont définies au voisinage d'un point x , et sont continues en ce point, alors la matrice hessienne est symétrique.

Théorème V.6.3. (Schwarz)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} , et $x \in U$. On suppose que f admet des dérivées partielles d'ordre 2 dans un voisinage de x , et que ces dérivées partielles sont *continues* en x . Alors la matrice hessienne en x est *symétrique*, i.e.

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Démonstration. On considère une fonction de 2 variables seulement (on peut se ramener à ce cas-là en gelant $n - 2$ variables). L'idée est d'écrire de deux manières la quantité

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2),$$

en suivant deux chemins différents entre (x_1, x_2) et $(x_1 + h_1, x_2 + h_2)$. On a en premier lieu

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) = f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) + f(x_1 + h_1, x_2) - f(x_1, x_2).$$

Les 2 derniers termes s'écrivent, d'après la formule de Taylor-Lagrange,

$$f(x_1 + h_1, x_2) - f(x_1, x_2) = h_1 \partial_1 f(x_1, x_2) + \frac{h_1^2}{2} \partial_{11} f(x_1 + \theta_1 h_1, x_2),$$

avec $\theta_1 \in]0, 1[$. La première différence du membre de droite s'écrit elle

$$f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) = h_2 \partial_2 f(x_1 + h_1, x_2) + \frac{h_2^2}{2} \partial_{22} f(x_1 + h_1, x_2 + \theta'_2 h_2)$$

Si l'on écrit maintenant la même quantité $f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2)$ de la façon suivante

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) = f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) + f(x_1, x_2 + h_2) - f(x_1, x_2),$$

que l'on utilise des développements de Taylor-Lagrange comme précédemment, et que l'on identifie les deux écritures, on obtient

$$\begin{aligned} 0 &= h_2 h_1 \left(\frac{\partial_2 f(x_1 + h_1, x_2) - \partial_2 f(x_1, x_2)}{h_1} - \frac{\partial_2 f(x_1, x_2 + h_2) - \partial_2 f(x_1, x_2)}{h_2} \right) \\ &+ \frac{h_2^2}{2} (\partial_{22} f(x_1 + h_1, x_2 + \theta'_2 h_2) - \partial_{22} f(x_1, x_2 + \theta_2 h_2)) \\ &+ \frac{h_1^2}{2} (\partial_{11} f(x_1 + \theta'_1 h_1, x_2 + h_2) - \partial_{11} f(x_1 + \theta_1 h_1, x_2)). \end{aligned}$$

Si l'on prend maintenant h_1 et h_2 égaux à ε , et que l'on fait tendre ε vers 0, les deux derniers termes sont des $o(\varepsilon^2)$ par continuité de la dérivée seconde. Le premier terme doit donc lui-même être un $o(\varepsilon^2)$, ce qui impose que la quantité entre parenthèse converge vers 0 avec ε , d'où le résultat. \square

Définition V.6.4. (Continue différentiability)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} . On dit que f est deux fois continûment différentiable sur U , et l'on écrit $f \in C^2(U)$, si toutes les dérivées partielles d'ordre 2 de f existent et sont continues sur U , ce qui est équivalent à dire que f admet une matrice hessienne $H(x)$ en tout point x de U , et que la correspondance $x \mapsto H(x)$ est continue.

Remarque V.6.5. En toute rigueur (voir à la fin de la section pour plus de détail), mais au prix de certaines définitions abstraites que nous avons choisi d'écarter, nous devrions définir la différentielle seconde comme l'application différentielle de la différentielle : $d^2 f = d(df)$, c'est à dire comme une application linéaire de \mathbb{R}^n dans l'espace des applications linéaires de \mathbb{R}^n dans \mathbb{R} . Et ensuite dire que l'application est C^2 si cette correspondance est continue, indépendamment des dérivées partielles premières ou secondes afférentes à une base particulière. On peut néanmoins montrer, dans l'esprit de la proposition V.1.8 pour les différentielles d'ordre 1, que la continuité de toutes les dérivées partielles secondes implique le caractère C^2 . Il est donc licite de fonder la définition précédente sur la caractérisation basée sur les dérivées partielles.

Proposition V.6.6. (Développement limité du gradient)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U . On a alors (h est pris suffisamment petit pour que $x + h \in U$)

$$\nabla f(x + h) = \nabla f(x) + H(x) \cdot h + \varepsilon(h) \|h\|.$$

Démonstration. Pour tout $i = 1, \dots, N$, la fonction $y \mapsto \partial_i f(y)$ est continûment différentiable sur U , et l'on a

$$\begin{aligned} \partial_i f(x + h) &= \partial_i f(x) + \langle \nabla \partial_i f(x) | h \rangle + \varepsilon(h) \|h\| \\ &= \partial_i f(x) + \sum_{j=1}^N \partial_j \partial_i f(x) h_j + \varepsilon(h) \|h\| \\ &= \partial_i f(x) + H \cdot h + \varepsilon(h) \|h\|, \end{aligned}$$

qui est l'identité annoncée. \square

Proposition V.6.7. (Développement limité à l'ordre 2)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U . On a alors (h est pris suffisamment petit pour que $x + h \in U$)

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + \frac{1}{2} \langle h | H(x) \cdot h \rangle + \varepsilon(h) \|h\|^2.$$

Démonstration. On introduit la fonction

$$h \mapsto g(h) = f(x + h) - f(x) - \langle \nabla f(x) | h \rangle - \frac{1}{2} \langle h | H(x) \cdot h \rangle.$$

On a

$$\nabla g(h) = \nabla f(x + h) - \nabla f(x) - H(x) \cdot h = \|h\| \varepsilon(h)$$

d'après la proposition V.6.6. Pour tout $\epsilon > 0$, il existe donc $\eta > 0$ tel que, pour tout h tel que $\|h\| \leq \eta$,

$$\|\nabla g(h)\| \leq \epsilon \|h\|.$$

On applique à présent le théorème des accroissements finis V.1.20 :

$$\|g(h)\| = \|g(h) - g(0)\| \leq \sup_{h' \in [0, h]} \|\nabla g(x + h')\| \|h\| \leq \epsilon \|h\|^2,$$

avec $g(h) = f(x + h) - f(x) - \langle \nabla f(x) | h \rangle - \frac{1}{2} \langle h | H(x) \cdot h \rangle$. \square

Proposition V.6.8. (Développement de Taylor avec reste intégral)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U , et h tel que le segment $[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$ soit inclus dans U . On a alors

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + \int_0^1 \langle H(x + th) \cdot h | h \rangle (1 - t) dt.$$

Démonstration. On pose $\Phi(t) = f(x + th)$. On a, par intégration par parties,

$$\begin{aligned} \Phi(1) &= \Phi(0) + \int_0^1 \Phi'(t) dt = \Phi(0) - [\Phi'(t)(1 - t)]_0^1 + \int_0^1 \Phi''(t)(1 - t) dt \\ &= \Phi(0) + \Phi'(0) + \int_0^1 \Phi''(t)(1 - t) dt. \end{aligned} \tag{V.6.1}$$

On a

$$\begin{aligned}\Phi(t) &= f(x + th), \\ \Phi(t + \varepsilon) &= f(x + th + t\varepsilon) = f(x + th) + \langle \nabla f(x + th) | h \rangle \varepsilon + o(\varepsilon),\end{aligned}$$

et donc

$$\Phi'(t) = \langle \nabla f(x + th) | h \rangle.$$

Enfin

$$\Phi'(t + \varepsilon) = \langle \nabla f(x + th + \varepsilon h) | h \rangle + \langle \nabla f(x + th) | h \rangle + \langle H(x + th) \cdot h | h \rangle \varepsilon + o(\varepsilon),$$

d'où

$$\Phi''(t) = \langle H(x + th) \cdot h | h \rangle.$$

On injecte ces expressions dans (V.6.1), ce qui donne la formule annoncée. \square

Définition V.6.9. (Laplacien)

Soit f une fonction définie sur un ouvert U de \mathbb{R}^n , à valeurs dans \mathbb{R} . On suppose que la matrice hessienne est définie en $x \in U$. On appelle laplacien de f en x , la quantité

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} = \text{tr}(H)$$

(trace de la hessienne de f). Pour les fonctions telles que cette quantité est définie sur U , on appelle *laplacien* cet opérateur noté Δ .

Cet opérateur s'écrit aussi $\Delta = \nabla \cdot \nabla$, où ∇ est le gradient, et $\nabla \cdot$ l'opérateur de divergence : pour tout champ de vecteur $u = (u_1, \dots, u_n)$,

$$\nabla \cdot u = \sum_{i=1}^n \frac{\partial u_i}{\partial x_i} = \text{tr} J,$$

où J est la matrice jacobienne de u vu comme application de \mathbb{R}^n dans \mathbb{R}^n .

V.6.2 Différentielles d'ordre supérieur pour les fonctions de \mathbb{R}^n dans \mathbb{R}^m

La notion de différentielle seconde découle de celle de la différentielle. Comme pour les fonction de \mathbb{R} dans \mathbb{R} , la différentielle seconde sera simplement la différentielle de la différentielle. Pour une fonction de départ de \mathbb{R} dans \mathbb{R}^m , cette différentielle est une application de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (définition V.1.4). Si nous souhaitons dériver cette différentielle, nous avons besoin d'une définition un peu plus générale, qui porte sur des applications à valeurs dans un espace vectoriel normé E .

Définition V.6.10. (Différentielle (\bullet))

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans un espace vectoriel normé E . On dit que f est différentiable en $x \in U$ s'il existe une application linéaire de \mathbb{R}^n dans E , notée $df(x)$, telle que

$$f(x + h) = f(x) + df(x) \cdot h + \varepsilon(h) \|h\| \quad (\text{V.6.2})$$

où $\varepsilon(h)$ est une application de \mathbb{R}^n dans E , telle que $\|\varepsilon(h)\|$ tend vers 0 quand h tend vers 0.

Définition V.6.11. (Différentielle seconde ($\bullet\bullet$))

Soit f une application différentiable dans un voisinage U d'un point $x \in \mathbb{R}^n$, à valeurs dans \mathbb{R}^m . On dit que f est deux fois différentiable en $x \in U$ si l'application $x \mapsto df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (muni de la norme d'opérateur canonique) est différentiable en x . La différentielle de df en x , notée $d^2f(x)$, est une application linéaire de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

On peut de la même manière, si df^2 est définie dans un voisinage de x , définir la différentielle d'ordre 3 par $d^3f = d(d^2f)$, qui est une application de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, et les différentielles d'ordre $k = 4, 5, \dots$

V.7 Exercices

Exercice V.7.1. Calculer les matrices hessiennes des applications suivantes

$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2), \quad g(x_1, x_2) = \sin(x_1 x_2).$$

Exercice V.7.2. Soit A une matrice carrée d'ordre n , et f l'application quadratique de \mathbb{R}^n dans \mathbb{R} définie par

$$x = (x_1, \dots, x_n) \mapsto f(x) = \langle A \cdot x \mid x \rangle.$$

- a) Calculer la matrice Hessienne de f .
- b) Dans quel cas cette matrice hessienne est-elle nulle ?

Exercice V.7.3. Soit f une fonction deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^n$, et h un vecteur de \mathbb{R}^n .

- a) Quelle est la limite de

$$\frac{f(x - \varepsilon h) - 2f(x) + f(x + \varepsilon h)}{\varepsilon^2}$$

quand ε tend vers 0 ?

- b) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U . On suppose de plus f convexe, c'est-à-dire telle que

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) \quad \forall x, y \in U, \quad \forall \theta \in]0, 1[.$$

Montrer que pour tout x de U , la matrice H est positive, c'est à dire que

$$\langle H(x) \cdot h \mid h \rangle \geq 0 \quad \forall h \in \mathbb{R}^n.$$

- c) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U . On suppose que f est λ -convexe, c'est-à-dire telle que

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) - \frac{\lambda}{2}\theta(1 - \theta)\|y - x\|^2 \quad \forall x, y \in U, \quad \forall \theta \in]0, 1[,$$

avec $\lambda \in \mathbb{R}$. Que peut-on en déduire sur la matrice $H(x)$, pour $x \in U$?

- d) (Condition suffisante d'optimalité locale)

On considère f une fonction deux fois continûment différentiable sur un ouvert U . On considère un point $x \in U$ en lequel le gradient de f s'annule, et tel que les valeurs propres de $H(x)$ sont toutes strictement positives. Que peut on dire de x vis-à-vis de f ?

- e) (Condition suffisante d'optimalité globale)

On suppose maintenant l'ouvert U convexe, et f convexe sur U . Montrer que x minimise f sur U , c'est à-dire-que

$$f(y) \geq f(x) \quad \forall y \in U.$$

- f) (Condition nécessaire d'optimalité)

On considère pour finir f une fonction deux fois continûment différentiable sur un ouvert U . On suppose que $x \in U$ est un minimiseur local de f . Montrer que $\nabla f(x) = 0$, et que $H(x)$ est une matrice positive.

Exercice V.7.4. Soit f une fonction deux fois continûment différentiable sur un ouvert U de \mathbb{R}^n , et telle que ∇f est de norme constante égale à un sur U . Montrer que

$$H(x) \cdot \nabla f(x) = 0.$$

pour tout x dans U .

Exercice V.7.5. On cherche ici à exprimer le fait que le laplacien quantifie l'écart entre la valeur ponctuelle d'une fonction et la moyenne des valeurs de la fonction au voisinage de ce point. En dimension 1, une telle propriété est donnée par le a) de l'exercice V.7.3. En dimension 2, cette propriété prend la forme exprimée ci-dessous.

Soit f une fonction à valeurs réelles deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^2$. On note e_θ le vecteur unitaire $(\cos \theta, \sin \theta)$. Montrer que

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \frac{1}{2\pi} \int_0^{2\pi} (f(x + \varepsilon e_\theta) - f(x)) d\theta = \frac{1}{4} \Delta f(x).$$

Exercice V.7.6. (Fonctions holomorphes / harmoniques)

On se place dans le cadre de l'exercice V.2.4, page 91. On écrit f , fonction de l'ouvert $\Omega \subset \mathbb{C}$ dans \mathbb{C} , sous la forme

$$f(x + iy) = P(x, y) + iQ(x, y)$$

où P et Q sont à valeurs réelles. Montrer que, si f est holomorphe, alors P et Q vérifient

$$\partial_x P - \partial_y Q = \partial_y P + \partial_x Q = 0.$$

En déduire que P et Q sont *harmoniques*, i.e. de laplacien nul :

$$\Delta Q = \Delta P = \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2} \right) P = 0.$$

N.B. on pourra admettre que P et Q sont deux fois continûment différentiables (il s'agit d'une conséquence du caractère holomorphe de f).

