

## II.3 Théorèmes des fonctions implicites et d'inversion locale

Le résultat principal de cette section est le théorème dit *des fonctions implicites*, très utiles dans de multiples contextes, 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 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 II.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 point<sup>8</sup> 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  $U \in \mathbb{R}^n$  et  $V \in \mathbb{R}^m$  de  $x_0$  et  $y_0$ , respectivement, et une fonction  $\Psi$  de  $U$  dans  $V$ , tels que

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

La fonction  $\Psi$  est continûment différentiable sur  $U$ , et sa différentielle s'exprime

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

*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

---

8. Comme précisé dans la remarque II.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.

plus loin), et l'on cherche  $y$  tel que  $f(x, y) = 0$ . Le processus itératif découle des considérations suivantes : 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  $\Phi_x$  est définie par

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

pour tout  $y$  tel que  $(x, y) \in W$ . Noter que  $y$  est point fixe de  $\Phi_x$  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  $x \in \overline{B}(x_0, r)$ , tout  $y \in \overline{B}(y_0, r)$  (on prend  $r$  suffisamment petit pour que  $\overline{B}(x_0, r) \times \overline{B}(y_0, r) \subset W$ ),

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

de telle sorte que

$$\forall x \in \overline{B}(x_0, r), y \in \overline{B}(y_0, r), \|d\Phi_x(y)\| \leq \kappa.$$

On a donc, pour tous  $y, y'$  dans  $\overline{B}(y_0, r)$ ,

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

d'après le théorème des accroissements finis, avec  $\kappa = 1/2$ . L'application  $\Phi_x$  est donc contractante sur  $\overline{B}(y_0, r)$ . Montrons qu'elle laisse stable une boule autour de  $y_0$ . Comme l'application

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

est continue en  $x_0$ , il existe un  $r' < r$  tel que, pour tout  $x \in \overline{B}(x_0, r')$ , 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  $x \in \overline{B}(x_0, r')$ , tout  $y \in \overline{B}(y_0, r)$ ,

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

Pout tout  $x \in \overline{B}(x_0, r')$ , l'application  $\Phi_x$  est donc bien définie de  $\overline{B}(y_0, r)$  dans lui-même, et cette 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  $\overline{B}(y_0, r)$ , c'est-à-dire qu'il existe un unique  $y \in \overline{B}(y_0, r)$  tel que  $f(x, y) = 0$ . On note  $\Psi$  l'application qui à  $x$  associe cette unique solution en  $y$  de  $f(x, y) = 0$ .

Montrons maintenant la continuité de  $\Psi$ , et précisons le choix des voisinages  $U$  et  $V$ . Soient  $x_1$  et  $x_2$  deux points de  $\overline{B}(x_0, r')$ , 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  $\Phi_{x_2}$  est  $\kappa$ -contractante sur  $\overline{B}(y_0, r)$ , on a  $\|\Phi_{x_2}(y_2) - \Phi_{x_2}(y_1)\| \leq \kappa \|y_2 - y_1\|$ , d'où

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

d'après le théorème des accroissements finis II.1.18 ( $f$  étant continûment différentiable sur le compact  $\overline{B}(x_0, r') \times \overline{B}(y_0, r)$ , sa différentielle partielle par rapport à  $x$  est bornée). Cette quantité tend en particulier vers 0 quand  $x_2$  tend vers  $x_1$ . L'application  $\Psi$  est donc continue sur  $\overline{B}(x_0, r')$  à valeurs dans  $\overline{B}(y_0, r)$ . Soit  $V$  voisinage ouvert de  $y_0$  inclus dans  $\overline{B}(y_0, r)$ . Comme  $\Psi$  est continue, il existe un voisinage ouvert de  $x_0$ ,  $U \subset \overline{B}(x_0, r')$ , tel que  $\Psi(U) \subset V$ .

Il reste à montrer que  $\Psi$  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). L'application  $x \mapsto \Psi(x)$  est donc 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 continument différentiable, et que  $\Psi$  est continue,  $x \mapsto d\Psi(x)$  est continue.  $\square$

**Remarque II.3.2.** On notera que, par construction,  $\Psi$  est bien définie sur tout  $U$ , mais 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).

**Remarque II.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 II.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*<sup>9</sup>. 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  $\varepsilon$  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 II.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  $\Psi$  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  $\Psi$  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

9. 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.

sera important de bien en maîtriser la valeur, alors que les paramètres pour lesquels  $\partial y_i / \partial x_j$  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 II.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 solution<sup>10</sup>.

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

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

$$df^{-1}(y) = (df(x))^{-1}.$$

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

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

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

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

qui conclut la preuve. □

**Théorème II.3.9.** (Inversion locale)

Soit  $f$  une application continûment différentiable d'un ouvert  $W \subset \mathbb{R}^n$  dans  $\mathbb{R}^n$ . On suppose que  $df(x)$  est inversible pour tout  $x \in W$ . Alors  $f$  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  $y_0 = f(x_0)$  tel que  $f|_U$  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  $y$ ) :

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

---

10. 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.

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

$$\partial_x g(y, x) = df(x).$$

Cette différentielle est inversible sur  $W$  par hypothèse. Soit  $x_0 \in W$ , et  $y_0 = f(x_0)$ , 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  $\tilde{U}$  de  $x_0$ , et  $\Psi$  une application continûment différentiable de  $V$  dans  $\tilde{U}$ , tels que

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

L'application  $\Psi$  est donc la réciproque de  $f$ . 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 II.3.2) :  $\Psi$ , 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  $y = f(x)$  n'a qu'une solution en  $x \in \tilde{U}$ , cet ensemble s'écrit aussi  $U = \tilde{U} \cap f^{-1}(V)$ . Il s'agit donc bien d'un ouvert par continuité de  $f$ .  $\square$

## II.4 Exercices

**Exercice II.4.1.** 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 II.4.2.** Identifier dans les cas suivantes 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.

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 II.4.3.** (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  $\Psi$  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  $\Psi$ .

**Exercice II.4.4.** (Dépendance d'une racine simple d'un polynôme complexe vis-à-vis des coefficients (version complexe de l'exercice II.4.3))

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

$$P_c(X) = c_0 + c_1X + \dots + c_nX^n.$$

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  $\Psi$  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 II.4.5.** 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,  $F_0 \approx 341 \text{ Wm}^{-2}$ . 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  $\sigma T^4$ , 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 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 valide<sup>11</sup>).

b) Montrer<sup>12</sup> 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  $\delta T$  autour de  $T_0$ , que peut-on dire (toujours dans l'hypothèse où l'on accorde une foi absolue au modèle) des variations  $\delta S$ ,  $\delta A$ , et  $\delta F$  qui ont pu induire cette variation de température?

c) On estime que le  $\text{CO}_2$  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  $\text{CO}_2$  est proportionnel à sa concentration dans l'atmosphère, estimer l'augmentation du taux de  $\text{CO}_2$  qui conduirait, selon ce modèle à une augmentation de la température de 2 °C.

---

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

12. 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.

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).



## II.5 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.

### II.5.1 Dérivées partielles d'ordre supérieur pour les fonctions scalaires

**Définition II.5.1.** (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 II.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 II.5.2.** (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

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

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 II.5.3.** (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 ses 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

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

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

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

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  $\varepsilon$ , et que l'on fait tendre  $\varepsilon$  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  $\varepsilon$ , d'où le résultat.  $\square$

**Définition II.5.4.** (Continue différentiabilité)

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 II.5.5.** 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 II.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 II.5.6.** (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  $U$ , et  $h$  tel que le segment  $[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$  soit inclus dans  $U$ . On a alors

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

*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 II.5.7.** (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  $U$ , et  $h$  tel que le segment  $[x, x+h] = \{x + \theta h, \theta \in [0, 1]\}$  soit inclus dans  $U$ . On a alors

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

*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

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

D'après la proposition II.5.6 ci-dessus, cette quantité est un  $o(h)$ . Pour tout  $\varepsilon > 0$ , il existe donc  $\eta > 0$  tel que, pour tout  $h$  tel que  $\|h\| \leq \eta$ ,

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

On applique à présent le théorème des accroissements finis II.1.18 :

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

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 II.5.8.** (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,

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

On a

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

qui conduit à la formule annoncée.  $\square$

**Définition II.5.9.** (Laplacien, opérateur 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 opérateur laplacien l'application qui à la fonction  $f$  associe la fonction  $\Delta f$ .

### II.5.2 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 II.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 II.5.10.** (Différentielle (•))

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

$$f(x+h) = f(x) + df(x) \cdot h + \varepsilon(h) \|h\| \quad (\text{II.5.1})$$

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 II.5.11.** (Différentielle seconde (•••))

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  $k = 4, 5, \dots$

## II.6 Exercices

**Exercice II.6.1.** Calculer les matrices hessiennes des applications suivantes

$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2), \quad f(x, y) = x_1^p x_2^q.$$

**Exercice II.6.2.** 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 II.6.3.** 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  $\varepsilon$  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  $\lambda$ -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 II.6.4.** Soit  $f$  une fonction deux fois continûment différentiable sur un ouvert  $U$  de  $\mathbb{R}^n$ , et telle que  $\nabla f$  est de norme constante égale à un sur  $U$ . Montrer que

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

pour tout  $x$  dans  $U$ .

**Exercice II.6.5.** 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 II.6.3. 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  $e_\theta$  le vecteur unitaire  $(\cos \theta, \sin \theta)$ . Montrer que

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