Mathématiques II
HEC Lausanne, Spring 2025

Theory sheet 10

Différentielle d'ordre deux

La dernière fois, nous avons vu comment approximer $f(\boldsymbol{x}+\Delta\boldsymbol{x})$ pour un petit $\Delta\boldsymbol{x}$ en utilisant la différentielle d'ordre un :

$$f(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx f(\boldsymbol{x}) + df(\boldsymbol{x},\Delta\boldsymbol{x}),$$

où

$$df(\boldsymbol{x},\Delta\boldsymbol{x}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\,\Delta x_j = (\operatorname{grad} f(\boldsymbol{x}))^\top \Delta\boldsymbol{x}$$

Rappelons que si l'on fait un pas $\boldsymbol{x} \rightsquigarrow \Delta\boldsymbol{x}$ dans la direction orthogonale à $\operatorname{grad} f(\boldsymbol{x})$, l'approximation ci-dessus dégénère :

$$f(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx f(\boldsymbol{x}),$$

ce qui est trivial. Si l'on souhaite tout de même capturer l'effet de ce déplacement sur f, il nous faut une information plus précise, fournie par l'objet suivant :

Définition 1. *La* <u>*différentielle d'ordre deux*</u> *d'une fonction* $f:\mathbb{R}^n \to \mathbb{R}$ *est l'objet* <u>*formel*</u> *suivant :*

$$d^2 f(\boldsymbol{x}) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i\,\partial x_j}\,dx_i\,dx_j.$$

Comme pour la différentielle d'ordre un, il faut voir $d^2 f$ comme une fonction de $\boldsymbol{x}$ et des accroissements formels dx_j, $j=1,\dots,n$. On peut alors <u>évaluer</u> $d^2 f(\boldsymbol{x})$ sur n'importe quel vecteur d'accroissements $\Delta\boldsymbol{x}$ de la façon suivante :

$$d^2 f(\boldsymbol{x},\Delta\boldsymbol{x}) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i\,\partial x_j}\,\Delta x_i\,\Delta x_j.$$

Dans le cas $n=2$, cela s'écrit :

$$d^2 f(\boldsymbol{x},\Delta\boldsymbol{x}) = f''_{x_1x_1}(\boldsymbol{x})\,(\Delta x_1)^2 + 2f''_{x_1x_2}(\boldsymbol{x})\,\Delta x_1\,\Delta x_2 + f''_{x_2x_2}(\boldsymbol{x})\,(\Delta x_2)^2.$$

Si $\Delta\boldsymbol{x}$ est petit, on a l'approximation suivante :

$$\boxed{f(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx f(\boldsymbol{x}) + df(\boldsymbol{x},\Delta\boldsymbol{x}) + \frac{1}{2}\,d^2 f(\boldsymbol{x},\Delta\boldsymbol{x})}$$

Si $n=2$, cela s'écrit :

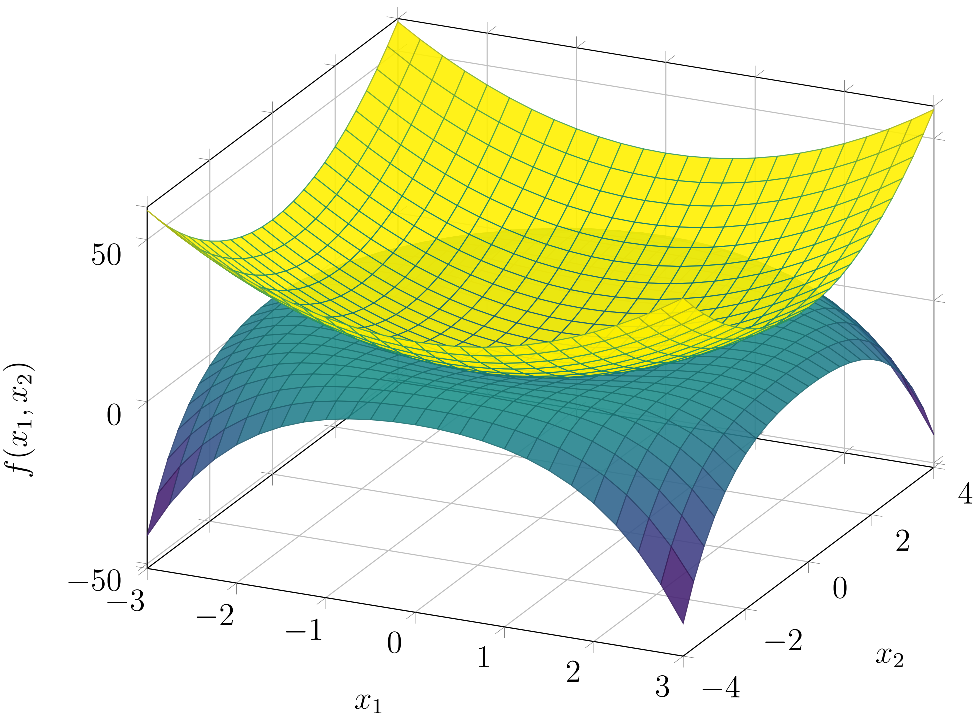
$$\begin{aligned} f(\boldsymbol{x}+\Delta\boldsymbol{x}) &\approx f(\boldsymbol{x}) + f'_{x_1}(\boldsymbol{x})\,\Delta x_1 + f'_{x_2}(\boldsymbol{x})\,\Delta x_2 \\ &\qquad + \frac{1}{2}\,f''_{x_1}(\boldsymbol{x})\,(\Delta x_1)^2 + f''_{x_1x_2}(\boldsymbol{x})\,\Delta x_1\,\Delta x_2 + \frac{1}{2}\,f''_{x_2}(\boldsymbol{x})\,(\Delta x_2)^2 \end{aligned}$$

Si $n=1$, cela devient encore plus simple :

$$f(x+\Delta) \approx f(x) + f'(x)\,\Delta + \frac{1}{2}\,f''(x)\,\Delta^2.$$

Voici quelques remarques :

- Si $\Delta \boldsymbol{x}$ est orthogonale à $\text{grad}\, f(\boldsymbol{x})$, on a $f(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx f(\boldsymbol{x}) + \frac{1}{2}\, d^2 f(\boldsymbol{x}, \Delta\boldsymbol{x})$, donc la dépendance en $\Delta\boldsymbol{x}$ ne disparaît plus !
- Pour d'autres $\Delta\boldsymbol{x}$, l'approximation est maintenant plus précise. On dit que la formule encadrée ci-dessus donne l'approximation d'ordre deux de f au voisinage de $\boldsymbol{x}$.
- À noter qu'en tant que fonction de $\Delta\boldsymbol{x}$, la différentielle d'ordre deux $d^2 f(\boldsymbol{x}, \Delta\boldsymbol{x})$ est une forme quadratique.
- Interprétation géométrique : si l'approximation d'ordre un correspond à la recherche de la meilleure *droite* ou plan épousant le paysage de f en un point donné, l'approximation d'ordre deux donne la meilleure approximation par paraboloïde/hyperboloïde de f au voisinage d'un certain $\boldsymbol{x}$. Voici une illustration de cela :



La surface jaune sur ce graphique est donnée par $z = 3x^2 + 2y^2$. Il s'agit d'une approximation d'ordre deux de $z = (3x^2 + 2y^2)(1 - x^2/10 - y^2/20)$. Remarquez comment les deux surfaces se touchent en $x = y = 0$, comment elles restent proches lorsque $(x, y) \approx (0, 0)$, mais elles divergent rapidement l'une de l'autre si l'on s'éloigne trop de $x = y = 0$.

Hessian

Définition 2. *Le Hessien d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ en un point $\boldsymbol{x} \in \mathbb{R}^n$ est la matrice $H(\boldsymbol{x}) \in M_{n,n}$ des dérivées partielles secondes :*

$$(H(\boldsymbol{x}))_{ij} = \frac{\partial^2 f}{\partial x_i\, \partial x_j}.$$

De la même manière que nous avons utilisé le gradient pour représenter la différentielle d'ordre un par

$$df(\boldsymbol{x}) = (\operatorname{grad} f(\boldsymbol{x}))^\top d\boldsymbol{x},$$

on peut utiliser le Hessien pour représenter la différentielle d'ordre deux :

$$\boxed{d^2 f(\boldsymbol{x}) = (d\boldsymbol{x})^\top H(\boldsymbol{x})\, d\boldsymbol{x}.}$$

Cette formule met encore plus en évidence le fait que $d^2 f(\boldsymbol{x})$ est une forme quadratique en $d\boldsymbol{x}$, comme mentionné ci-dessus.

Notez que $H(\boldsymbol{x})$ est symétrique, car les dérivées partielles commutent :

$$(H(\boldsymbol{x}))_{ij} = \frac{\partial^2 f}{\partial x_i\, \partial x_j} = \frac{\partial^2 f}{\partial x_j\, \partial x_i} = (H(\boldsymbol{x}))_{ij}.$$

En combinant la description de $df(\boldsymbol{x})$ en termes de $\operatorname{grad} f(\boldsymbol{x})$ et celle de $d^2 f(\boldsymbol{x})$ en termes de $H(\boldsymbol{x})$, on obtient

$$\boxed{f(\boldsymbol{x} + \Delta\boldsymbol{x}) \approx f(\boldsymbol{x}) + (\operatorname{grad} f(\boldsymbol{x}))^\top \Delta\boldsymbol{x} + \frac{1}{2}\,(\Delta\boldsymbol{x})^\top H(\boldsymbol{x})\, \Delta\boldsymbol{x}.}$$

Cette formule n'est qu'une autre représentation de l'approximation d'ordre deux de f. De telles approximations sont appelées développements de Taylor (d'ordre un/deux). Il existe aussi des développements de Taylor d'ordre supérieur (utilisant des dérivées d'ordre plus élevé).

Exemple

Prenons $n = 2$, considérons $f(x, y) = x^y$ définie pour $x, y > 0$. Cherchons son développement de Taylor au voisinage de $x = 1$ et $y = 2$:

$$\begin{aligned}
f(1,2) &= 1^2 = 1 \\
f'_x(1,2) &= yx^{y-1} \,|_{x=1,y=2} = 2 \\
f'_y(1,2) &= x^y \ln x \,|_{x=1,y=2} = 0 \\
f''_{xx}(1,2) &= y(y-1)x^{y-2} \,|_{x=1,y=2} = 2 \\
f''_{xy}(1,2) &= 1 \cdot x^{y-1} + yx^{y-1} \ln x \,|_{x=1,y=2} = 1 \\
f''_{yy}(1,2) &= x^y (\ln x)^2 \,|_{x=1,y=2} = 0.
\end{aligned}$$

Donc,

$$\operatorname{grad} f(1,2) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad H(1,2) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}.$$

On obtient donc la formule de Taylor d'ordre deux suivante :

$$f(1 + \Delta x, 2 + \Delta y) \approx 1 + 2 \cdot \Delta x + 0 \cdot \Delta y + \frac{1}{2} \cdot 2(\Delta x)^2 + 1 \cdot \Delta x\, \Delta y + \frac{1}{2} \cdot 0 \cdot (\Delta y)^2.$$

On peut aussi l'écrire ainsi :

$$f(x, y) = 1 + 2 \cdot (x - 1) + 0 \cdot (y - 2) + \frac{1}{2} \cdot 2(x-1)^2 + 1 \cdot (x-1)\,(y-2) + \frac{1}{2} \cdot 0 \cdot (y-2)^2,$$

où $x = 1 + \Delta x \implies \Delta x = x - 1$ et $y = 2 + \Delta y \implies \Delta y = y - 2$.

Par exemple, notre approximation donne

$$f(1.01, 2.01) \approx 1.0202,$$

alors que la valeur exacte est

$$f(1.01, 2.01) = 1.020201508\dots$$

Extrema libres (non contraints) : conditions du premier ordre

Définition 3. *Un point $\boldsymbol{x}_0 \in \mathbb{R}^n$ est un point de maximum local d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ si $f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)$ pour tout $\boldsymbol{x}$ suffisamment proche de $\boldsymbol{x}_0$.*

C'est un point de minimum local si $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$ pour tout $\boldsymbol{x}$ suffisamment proche de $\boldsymbol{x}_0$.

Rappelons que pour trouver les extrema locaux d'une fonction lisse d'une variable $f : \mathbb{R} \to \mathbb{R}$, on commence par chercher les points candidats à l'extremum, c'est-à-dire les x tels que

$$f'(x) = 0.$$

Certains de ces points peuvent ne pas être extrémaux (il faut vérifier les conditions du second ordre), mais si x est un extremum local, alors $f'(x) = 0$ (la condition est nécessaire).

Si $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction de plusieurs variables et que $\boldsymbol{x}$ est son minimum ou maximum, alors en particulier c'est un extremum d'une fonction d'une variable $g(x_j) = f(\boldsymbol{x})$ (toutes les autres variables sont fixées sauf une ; assurez-vous de bien comprendre cet argument !). Donc $g'(x_j) = 0$, ou

$$g'(x_j) = \frac{\partial f}{\partial x_j} = 0.$$

Ainsi, toutes les dérivées partielles doivent être nulles en un point extrémal. On peut écrire cela de façon concise :

$$\boxed{f'_{x_j}(\boldsymbol{x}) = 0 \quad \text{pour tout } j = 1, \dots, n,}$$

ou en utilisant la notation du gradient :

$$\boxed{\operatorname{grad} f(\boldsymbol{x}) = \boldsymbol{0}.}$$

Formulons cela comme un théorème :

Théorème 1 (Conditions du premier ordre)**.** *Si $\boldsymbol{x}$ est un point de minimum local ou de maximum local de $f : \mathbb{R}^n \to \mathbb{R}$, alors*

$$\operatorname{grad} f(\boldsymbol{x}) = \boldsymbol{0}.$$

Comme dans le cas univarié, il s'agit d'une condition nécessaire pour que $\boldsymbol{x}$ soit un extremum, *mais pas suffisante.*

Remarque 1. *Considérons $f(x) = x^3$. Clairement, $f'(0) = 3x^2 \mid_{x=0} = 0$, mais $x = 0$ n'est ni un minimum, ni un maximum de f.*

Type d'extremum : conditions du second ordre

Si on a trouvé $\boldsymbol{x}_0$ tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$, alors on a l'approximation suivante de f au voisinage de ce point :

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^\top H(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0).$$

Comment savoir si $\boldsymbol{x}_0$ est un minimum ou un maximum ? Peut-il n'être ni l'un ni l'autre ?

Remarquons que $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$ pour $\boldsymbol{x} \approx \boldsymbol{x}_0$ si

$$(\boldsymbol{x} - \boldsymbol{x}_0)^\top H(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0.$$

Si $\boldsymbol{x}$ est un minimum local, cette condition doit être vérifiée pour tout $\boldsymbol{x} \approx \boldsymbol{x}_0$, ce qui par définition signifie que $H(\boldsymbol{x}_0)$ est semi-définie positive.

De même, si $\boldsymbol{x}$ est un point de maximum local, alors $f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)$ pour tout $\boldsymbol{x} \approx \boldsymbol{x}_0$, ce qui est équivalent à $H(\boldsymbol{x}_0)$ étant semi-définie négative.

Ce sont encore des conditions nécessaires, mais qu'en est-il des conditions suffisantes ? Plus précisément, si on a trouvé $\boldsymbol{x}_0$ tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$ et $H(\boldsymbol{x}_0)$ est semi-définie positive, peut-on conclure que f admet un minimum en ce point ? La réponse est : oui, si $H(\boldsymbol{x}_0)$ est définie positive (strictement).

Théorème 2. *Si $\boldsymbol{x}_0$ est un point tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$ et $H(\boldsymbol{x}_0)$ est définie positive, alors f admet un minimum local en ce point.*

Si $\boldsymbol{x}_0$ est un point tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$ et $H(\boldsymbol{x}_0)$ est définie négative, alors f admet un maximum local en ce point.

Si $\boldsymbol{x}_0$ est un point tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$ et $H(\boldsymbol{x}_0)$ est indéfinie, alors f n'admet ni maximum local, ni minimum local en ce point.

Si $\boldsymbol{x}_0$ est un point tel que $\operatorname{grad} f(\boldsymbol{x}_0) = \mathbf{0}$ et $H(\boldsymbol{x}_0)$ est semi-définie (positive ou négative), une analyse supplémentaire est nécessaire pour conclure s'il s'agit d'un maximum, d'un minimum ou d'aucun des deux.

Remarque 2. *Si $H(\boldsymbol{x}_0)$ est seulement semi-définie, vérifier si $\boldsymbol{x}_0$ est un extremum local demande une analyse plus poussée. Reprenons la remarque précédente : si $f(x) = x^3$, son Hessien est une matrice 1×1 identifiée à sa dérivée seconde : $H(x) = f''(x) = 6x$. En zéro, ce Hessien est nul : $H(0) = 0$, donc à la fois ≥ 0 et ≤ 0. Cependant, $x_0 = 0$ n'est clairement pas un extremum local de f.*

Considérons un autre exemple : $f(x) = x^4$. Dans ce cas $H(0) = 0$, mais on a bien un minimum local en $x_0 = 0$. Ces exemples montrent qu'il faut des développements plus précis pour vérifier l'existence d'un extremum en un point où le Hessien est semi-défini.

Remarque 3. *Rappelons que dire que $H(\boldsymbol{x}_0)$ est d'un certain type (définie positive (semi-)/définie négative (semi-)/indéfinie) revient à dire que la forme quadratique correspondante $d^2f(\boldsymbol{x}_0)$ est de ce type. C'est pourquoi on parle souvent de $d^2f(\boldsymbol{x}_0)$ comme étant d'un certain type.*

Remarque 4. *Géométriquement, un Hessien définie positive (négative) signifie que la fonction f ressemble à un paraboloïde ouvert vers le haut (bas) au voisinage de $\boldsymbol{x}_0$. Si $H(\boldsymbol{x}_0)$ est indéfinie, f ressemble à un hyperboloïde près de $\boldsymbol{x}_0$. Dans ce cas, on dit que f admet un point selle en $\boldsymbol{x}_0$.*

Example

Soit $f(x,y) = xe^{-x^2-y^2}$. Alors

$$f'_x = 0 \implies e^{-x^2-y^2} - 2x^2e^{-x^2-y^2} = 0 \implies x = \pm\frac{1}{\sqrt{2}}$$

et de même

$$f'_y = 0 \implies -2ye^{-x^2-y^2} = 0 \implies y = 0.$$

On a donc deux points candidats à l'extremum : $(\frac{1}{\sqrt{2}}, 0)$ et $(-\frac{1}{\sqrt{2}}, 0)$. Vérifions leur nature :

$$H(\boldsymbol{x}) = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{xy} & f''_{yy} \end{pmatrix} = e^{-x^2-y^2}\begin{pmatrix} 2x(2x^2-3) & 2y(2x^2-1) \\ 2y(2x^2-1) & 2x(2y^2-1) \end{pmatrix}.$$

Au point $(\frac{1}{\sqrt{2}}, 0)$ on a

$$H\left(\frac{1}{\sqrt{2}}, 0\right) = e^{-\frac{1}{2}}\begin{pmatrix} \frac{-4}{\sqrt{2}} & 0 \\ 0 & -\frac{2}{\sqrt{2}} \end{pmatrix}.$$

Cette matrice est définie négative, donc $(\frac{1}{\sqrt{2}}, 0)$ est un maximum local. Ensuite, au point $(-\frac{1}{\sqrt{2}}, 0)$ on a

$$H\left(-\frac{1}{\sqrt{2}}, 0\right) = e^{-\frac{1}{2}}\begin{pmatrix} \frac{4}{\sqrt{2}} & 0 \\ 0 & \frac{2}{\sqrt{2}} \end{pmatrix}.$$

Comme cette matrice est définie positive, $(-\frac{1}{\sqrt{2}}, 0)$ est un minimum local de f.