

Optimisation

Chapitre 3 : Différentiabilité, Convexité

Joseph GERGAUD, Serge GRATTON & Daniel RUIZ

15 septembre 2023

Le but de ce chapitre 3 est d'étudier les **les notions de dérivée et de convexité**

Définition 3.1.1 – Dérivée d'une fonction de \mathbb{R} dans \mathbb{R}

Une fonction d'une seule variable réelle à valeurs dans \mathbb{R} est dérivable en un point x de \mathbb{R} s'il existe un nombre réel a noté $f'(x)$ tel que :

$$\lim_{t \rightarrow 0} \frac{f(x+t) - f(x) - at}{t} = 0 .$$

Définition 3.1.2 – Dérivée au sens de Fréchet

Soient E et F deux espaces vectoriels normés. Soit f une application définie sur le domaine $D \subset E$ et à valeurs dans F . L'application f est dite **F-différentiable** (ou différentiable au sens de Fréchet, ou encore différentiable au sens fort) en un point x de l'intérieur du domaine D , s'il existe un opérateur linéaire continu $f'(x)$ de E dans F ($f'(x) \in \mathcal{L}(E, F)$), tel que

$$\forall \mathbf{h} \in E , \quad f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + f'(\mathbf{x}) \cdot \mathbf{h} + \|\mathbf{h}\|_E \varepsilon(\mathbf{h}) , \quad \text{avec} \quad \lim_{\|\mathbf{h}\|_E \rightarrow 0} \|\varepsilon(\mathbf{h})\|_F = 0 . \quad (1)$$

Si $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$ les bases étant choisies dans ces espaces, on peut associer à l'application linéaire $f'(\mathbf{x})$ une matrice.

Définition 3.1.3 – Matrice jacobienne

Une base dans les espaces de départ et d'arrivée étant choisie, on appelle matrice jacobienne la matrice vérifiant

$$\forall \mathbf{h} \in \mathbb{R}^n, \quad f'(\mathbf{x}).\mathbf{h} = J_f(\mathbf{x}) \times \mathbf{h}.$$

Proposition 3.1.4

Si l'application f est F -différentiable (dérivable) au point x , elle est alors continue au point x .

Définition 3.1.5

Soient E et F deux espaces vectoriels normés, et $\Omega \subset E$ un ouvert de E . On dit que **l'application** $f : \Omega \subset E \rightarrow F$ **est dérivable dans** Ω si elle est dérivable en tout point x de Ω . On peut alors définir l'application

$$f' : x \in \Omega \subset E \rightarrow f'(x) \in \mathcal{L}(E, F) ,$$

appelée **application dérivée de** f . Si l'application dérivée $f' : \Omega \subset E \rightarrow \mathcal{L}(E, F)$ est continue, on dit que **l'application** $f : \Omega \subset E \rightarrow F$ **est (une fois) continûment dérivable dans** Ω , et on écrit

$$f \in \mathcal{C}^1(\Omega) .$$

- **Exemple 3.1.1.**

$$\begin{array}{lll} f: \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ x & \longmapsto & Ax + b \end{array}$$

- **Exemple 3.1.1.**

$$\begin{array}{lll} f: \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \\ x & \longmapsto & Ax + b \end{array}$$

- **Exemple 3.1.2.**

$$\begin{array}{lll} f: \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \frac{1}{2}x^T Ax + b^T x + c \end{array}$$

▪ Exemple 3.1.1.

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longmapsto Ax + b \end{aligned}$$

▪ Exemple 3.1.2.

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}x^T Ax + b^T x + c \end{aligned}$$

▪

Définition 3.1.6 – Gradient

On appelle gradient de la fonction f en \mathbf{x} de \mathbf{H} , espace de Hilbert, à valeur dans \mathbb{R} l'unique vecteur de \mathbf{H} , noté $\nabla f(\mathbf{x})$, tel que

$$\forall \mathbf{h} \in \mathbf{H}, \quad f'(\mathbf{x}).\mathbf{h} = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle$$

▪ Exemple 3.1.3.

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \beta &\longmapsto \frac{1}{2}\|y - X\beta\|^2 \end{aligned}$$

La donnée d'une application

$$f : \Omega \subset E \rightarrow F = \prod_{i=1}^p F_i$$

revient à se donner p applications composantes $f_i : \Omega \subset E \rightarrow F_i$, $1 \leq i \leq p$, de telle façon que

$$\forall \mathbf{x} \in E, \quad f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_p(\mathbf{x}) \end{pmatrix} .$$

Proposition 3.1.7

On établit facilement que **l'application f est dérivable en un point $\mathbf{a} \in \Omega$ si et seulement si chaque application composante l'est aussi**, et on a alors :

$$f'(\mathbf{a}) = \begin{pmatrix} f'_1(\mathbf{a}) \\ f'_2(\mathbf{a}) \\ \vdots \\ f'_p(\mathbf{a}) \end{pmatrix} \quad \text{avec } f'_i(\mathbf{a}) \in \mathcal{L}(E, F_i), \quad 1 \leq i \leq p .$$

Considérons ensuite une application

$$f : \Omega \subset \prod_{i=1}^n E_i \rightarrow F$$

où Ω est un ouvert de $\prod_{i=1}^n E_i$. Soit \mathbf{a} un point de Ω de composantes (a_1, a_2, \dots, a_n) , et soit $k \in (1, 2, \dots, n)$ l'un des indices.

Soit $\Omega_i, i = 1, \dots, n$ des ouverts des E_i tels que $\prod_{i=1}^n \Omega_i \subset \Omega$. On définit alors la k -ième **application partielle**

$$\begin{aligned} \Omega_k \subset E_k &\longrightarrow F \\ x_k &\longrightarrow f(a_1, \dots, a_{k-1}, x_k, a_{k+1}, \dots, a_n) \end{aligned} \quad .$$

Définition 3.1.8 – Dérivée partielle

On appelle dérivée partielle de f au point \mathbf{a} par rapport à la k -ième variable la dérivée, si elle existe de l'application partielle au point $a_k \in \Omega_k \subset E_k$, on note

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) \in \mathcal{L}(E_k, F)$$

cette dérivée partielle.

- Si f est une fonction de $\Omega \subset \mathbb{R}^n$, Ω ouvert à valeurs dans \mathbb{R}^m dérivable en $\mathbf{a} \in \Omega$ alors sa matrice jacobienne d'écrit

$$f'(\mathbf{a}) \cdot \mathbf{h} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{a}) & \frac{\partial f_p}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_p}{\partial x_n}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix}.$$

- Si $p = 1$ alors le gradient de f en \mathbf{a} s'écrit

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{pmatrix}.$$

- **Example 3.1.4.**

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2} \|x\|^2. \end{aligned}$$

- **Exemple 3.1.4.**

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2} \|x\|^2. \end{aligned}$$

- **Exemple 3.1.5.**

$$\begin{aligned} f: \mathbb{R}^3 &\longrightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) &\longmapsto \begin{pmatrix} x_2 \cos x_1 - x_3 \sin x_1 \\ x_1 x_2 x_3 \end{pmatrix}. \end{aligned}$$

Théorème 3.1.9 – Théorème des fonctions composées

Soient E , F , et G , trois espaces vectoriels normés. Soit $f : \Omega \subset E \rightarrow F$ une application dérivable en un point $\mathbf{x} \in \Omega$ (Ω ouvert de E), et soit $g : \tilde{\Omega} \subset F \rightarrow G$ une application dérivable au point $\mathbf{y} = f(\mathbf{x}) \in \tilde{\Omega}$ ($\tilde{\Omega}$ ouvert de F). On suppose $f(\Omega) \subset \tilde{\Omega}$. Alors l'application composée

$$g \circ f : \Omega \subset E \rightarrow G$$

est dérivable au point $\mathbf{x} \in \Omega$ et

$$\forall \mathbf{h} \in E, \quad (g \circ f)'(\mathbf{x}).\mathbf{h} = g'(f(\mathbf{x})).(f'(\mathbf{x}).\mathbf{h}).$$

Proposition 3.1.10 – Cas de la dimension finie

Si $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ et $G = \mathbb{R}^p$, on a

$$J_{g \circ f}(\mathbf{x}) = J_g(\mathbf{y}) \times J_f(\mathbf{x}) = J_g(f(\mathbf{x})) \times J_f(\mathbf{x}).$$

Exemple 3.1.6.

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \beta &\longmapsto \frac{1}{2} \|y - X\beta\|^2. \end{aligned}$$

Définition 3.1.11 – Dérivée seconde

Soit $f : \Omega \subset E \rightarrow F$ une application dérivable sur l'ouvert $\Omega \subset E$. Si l'application dérivée

$$f' : \Omega \subset E \rightarrow \mathcal{L}(E, F)$$

est elle-même dérivable (i.e. F -différentiable) en un point $\mathbf{x} \in \Omega$, sa dérivée, notée

$$f''(\mathbf{x}) \stackrel{\text{déf}}{=} (f')'(\mathbf{x}) \in \mathcal{L}(E, \mathcal{L}(E, F)) ,$$

est appelée **dérivée seconde de l'application f au point \mathbf{x}** , et on dit que **l'application f est deux fois dérivable au point \mathbf{x}** .

Notation : Il est facile de remarquer que l'application

$$\mathbf{B} : (\mathbf{h}, \mathbf{k}) \in E \times E \rightarrow ((f''(\mathbf{x}) \cdot \mathbf{h}) \cdot \mathbf{k}) \in F ,$$

est linéaire séparément en chacune des variables \mathbf{h} et \mathbf{k} , et est de ce fait **bilinéaire**. On identifie l'application dérivée seconde de f au point \mathbf{x} , $f''(\mathbf{x}) \in \mathcal{L}(E, \mathcal{L}(E, F))$, à une application de l'espace $\mathcal{L}_2(E \times E, F)$, espace des applications bilinéaires continues de $E \times E$ dans F . On écrira alors

$$\forall \mathbf{h}, \mathbf{k} \in E , \quad f''(\mathbf{x})(\mathbf{h}, \mathbf{k}) = (f''(\mathbf{x}) \cdot \mathbf{h}) \cdot \mathbf{k} .$$

Proposition 3.1.12

Si l'application $f : \Omega \subset E \rightarrow F$ est deux fois dérivable au point \mathbf{x} de l'ouvert $\Omega \subset E$, alors l'application dérivée seconde de f au point \mathbf{x} est une **application bilinéaire symétrique** en ce sens que

$$\forall \mathbf{h}, \mathbf{k} \in E, \quad f''(\mathbf{x})(\mathbf{h}, \mathbf{k}) = f''(\mathbf{x})(\mathbf{k}, \mathbf{h}).$$

Définition 3.1.13 – Application dérivée seconde

On dit que l'**application** $f : \Omega \subset E \rightarrow F$ est **deux fois dérivable dans** Ω si elle est deux fois dérivable en tout point \mathbf{x} de Ω . On peut alors définir l'**application dérivée seconde de** f

$$f'' : \mathbf{x} \in \Omega \subset E \rightarrow f''(\mathbf{x}) \in \mathcal{L}_2(E \times E, F) .$$

Si cette dernière application est continue, l'application f est dite **deux fois continûment dérivable dans** Ω , et on écrit

$$f \in \mathcal{C}^2(\Omega) .$$

Remarque 3.1.1. En ce qui concerne le **calcul** effectif des dérivées secondes, on utilise le résultat suivant, qui permet de se ramener à des calculs de dérivées premières : étant donné deux vecteurs quelconques $\mathbf{h}, \mathbf{k} \in E$, l'élément $f''(\mathbf{x})(\mathbf{h}, \mathbf{k}) \in F$ est égal à la dérivée au point $\mathbf{x} \in \Omega$ de l'application $\mathbf{v} \in \Omega \rightarrow f'(\mathbf{v}) \cdot \mathbf{k} \in F$, appliquée au vecteur \mathbf{h} .

Remarque 3.1.2. Cas où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ Si f est une application de \mathbb{R}^n à valeurs dans \mathbb{R} alors $f''(\mathbf{a})$ est une forme bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ à valeurs dans \mathbb{R} . On peut alors lui associer une matrice (n, n) appelée matrice hessienne

$$H_f(\mathbf{a}) = \nabla^2 f(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix},$$

et on a

$$\forall \mathbf{h} \in \mathbb{R}^n, \forall \mathbf{k} \in \mathbb{R}^n, f''(\mathbf{a}).(\mathbf{h}, \mathbf{k}) = \mathbf{h}^T \nabla^2 f(\mathbf{a}) \mathbf{k}.$$

- **Example 3.1.7.**

$$\begin{aligned} f: \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2} \|x\|^2. \end{aligned}$$

- **Exemple 3.1.7.**

$$\begin{array}{rcl} f: \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \frac{1}{2} \|x\|^2. \end{array}$$

- **Exemple 3.1.8.**

$$\begin{array}{rcl} f: \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \frac{1}{2} \|r(x)\|^2 \end{array}$$

(avec $r(x)$ deux fois dérivable).

Théorème 3.1.14 – Formules de Taylor pour les applications une fois dérivables

Soient $f : \Omega \subset E \rightarrow F$ et $[\mathbf{a}, \mathbf{a} + \mathbf{h}]$ un segment fermé contenu dans Ω ouvert.

i) Si f est dérivable en \mathbf{a} , alors

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \mathbf{h} + \|\mathbf{h}\|_E \varepsilon(\mathbf{h}) \quad , \quad \text{avec} \quad \lim_{\|\mathbf{h}\|_E \rightarrow 0} \|\varepsilon(\mathbf{h})\|_F = 0 \quad .$$

ii) **Formule des accroissements finis** : si $f \in \mathcal{C}^0(\Omega)$ et f est dérivable en tout point du segment ouvert $] \mathbf{a}, \mathbf{a} + \mathbf{h} [$, alors

$$\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a})\|_F \leq \sup_{\mathbf{x} \in] \mathbf{a}, \mathbf{a} + \mathbf{h} [} \|f'(\mathbf{x})\|_{\mathcal{L}(E, F)} \|\mathbf{h}\|_E \quad .$$

iii) **Formule de Taylor-Maclaurin** : si $f \in \mathcal{C}^0(\Omega)$ et f est dérivable en tout point du segment ouvert $] \mathbf{a}, \mathbf{a} + \mathbf{h} [$, et si $F = \mathbb{R}$, alors

$$\exists \theta \in]0, 1[\quad \text{tel que} \quad f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a} + \theta \mathbf{h}) \cdot \mathbf{h} \quad .$$

iv) **Formule de Taylor avec reste intégral** : si $f \in \mathcal{C}^1(\Omega)$ et F est complet, alors

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \int_0^1 (f'(\mathbf{a} + t\mathbf{h}) \cdot \mathbf{h}) dt \quad .$$

Théorème 3.1.15 – Formules de Taylor pour les appli. deux fois dérivables

Soit $f : \Omega \subset E \rightarrow F$ et $[\mathbf{a}, \mathbf{a} + \mathbf{h}]$ un segment fermé contenu dans Ω (Ω ouvert de E).

i) **Formule de Taylor-Young** : si f est dérivable dans Ω , et si f est deux fois dérivable au point \mathbf{a} , alors

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} f''(\mathbf{a})(\mathbf{h}, \mathbf{h}) + \|\mathbf{h}\|_E^2 \varepsilon(\mathbf{h}) \quad , \quad \text{avec} \quad \lim_{\|\mathbf{h}\|_E \rightarrow 0} \|\varepsilon(\mathbf{h})\|_F = 0 .$$

ii) **Formule des accroissements finis généralisée** : si $f \in \mathcal{C}^1(\Omega)$ et f est deux fois dérivable en tout point du segment ouvert $] \mathbf{a}, \mathbf{a} + \mathbf{h} [$, alors

$$\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - f'(\mathbf{a}) \cdot \mathbf{h}\|_F \leq \frac{1}{2} \sup_{\mathbf{x} \in] \mathbf{a}, \mathbf{a} + \mathbf{h} [} \|f''(\mathbf{x})\|_{\mathcal{L}_2(E \times E, F)} \|\mathbf{h}\|_E^2 .$$

iii) **Formule de Taylor-Maclaurin** : si $f \in \mathcal{C}^1(\Omega)$ et f est deux fois dérivable en tout point du segment ouvert $] \mathbf{a}, \mathbf{a} + \mathbf{h} [$, et si $F = \mathbb{R}$, alors

$$\exists \theta \in]0, 1[\quad \text{tel que} \quad f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} f''(\mathbf{a} + \theta \mathbf{h})(\mathbf{h}, \mathbf{h}) .$$

iv) **Formule de Taylor avec reste intégral** : si $f \in \mathcal{C}^2(\Omega)$ et F est un espace complet, alors

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \mathbf{h} + \int_0^1 (1-t)(f''(\mathbf{a} + t\mathbf{h})(\mathbf{h}, \mathbf{h})) dt .$$

Pour illustrer ces considérations, voici trois façons équivalentes d'écrire (par exemple) la formule de Taylor-Young pour une fonctionnelle $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois dérivable :

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} f''(\mathbf{a})(\mathbf{h}, \mathbf{h}) + \|\mathbf{h}\|_2^2 \varepsilon(\mathbf{h})$$

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle + \frac{1}{2} \langle \nabla^2 f(\mathbf{a}) \mathbf{h}, \mathbf{h} \rangle + \langle \mathbf{h}, \mathbf{h} \rangle \varepsilon(\mathbf{h})$$

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + (\nabla f(\mathbf{a}))^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T \nabla^2 f(\mathbf{a}) \mathbf{h} + \mathbf{h}^T \mathbf{h} \varepsilon(\mathbf{h})$$

$$f(\mathbf{a} + \mathbf{h}) = q_{\mathbf{a}}(\mathbf{h}) + \mathbf{h}^T \mathbf{h} \varepsilon(\mathbf{h}).$$

où $q_{\mathbf{a}}$ est une forme quadrique généralisée.

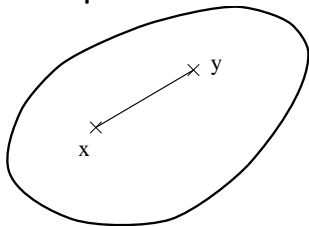
On indique dans ce paragraphe quelques propriétés de base d'une classe très importante de fonctionnelles.

Définition 3.2.1 – Ensembles convexes

L'ensemble D_0 est dit **convexe** si et seulement si

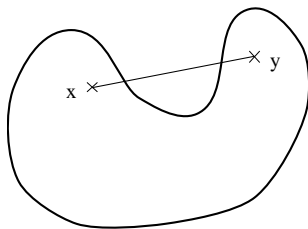
$$\forall \mathbf{x} \in D_0, \forall \mathbf{y} \in D_0, \forall \alpha \in [0, 1] \subset \mathbb{R} \text{ on a } \alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \in D_0 .$$

Remarque 3.2.1.



autrement dit, si $\mathbf{x} \in D_0$ et $\mathbf{y} \in D_0$, alors le segment qui joint ces deux points est également contenu dans D_0 , le segment $[\mathbf{x}, \mathbf{y}]$ étant défini par

$$\mathbf{z} \in [\mathbf{x}, \mathbf{y}] \iff \exists \alpha \in [0, 1] \text{ t.q. } \mathbf{z} = \alpha \mathbf{x} + (1 - \alpha) \mathbf{y} .$$

Exemple 3.2.1. Exemple d'ensemble non convexe

Remarque 3.2.2. la notion d'ensemble convexe correspond en fait à une propriété de régularité du domaine D_0 considéré

Définition 3.2.2 – Fonctionnelles convexes

Une fonctionnelle $f : D_0 \subset E \rightarrow \mathbb{R}$ est **convexe** sur le domaine convexe $D_0 \subset E$ (E espace vectoriel normé) si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \forall \alpha \in]0, 1[, \quad f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) .$$

La fonctionnelle f est **strictement convexe** sur le domaine convexe D_0 si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \mathbf{x} \neq \mathbf{y}, \forall \alpha \in]0, 1[, \quad f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) .$$

La fonctionnelle f est **uniformément convexe** sur le domaine convexe D_0 si il existe une constante $c > 0$ telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \forall \alpha \in]0, 1[,$$

$$\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) - f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \geq c \alpha (1 - \alpha) \|\mathbf{x} - \mathbf{y}\|_E^2 .$$

Remarque 3.2.3.

- i) Il est clair que la *convexité uniforme* entraîne la *convexité stricte* qui à son tour entraîne la *convexité*.
- ii) La convexité indique une certaine régularité de la fonctionnelle. En dimension finie, par exemple, la convexité peut induire des propriétés de continuité (c.f. proposition suivante).

Proposition 3.2.3

Soit $f : D_0 \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle convexe sur l'ouvert convexe $D_0 \subset \mathbb{R}^n$. Alors f est continue sur D_0 .

Théorème 3.2.4 – Caractérisation de la convexité à l'aide de la dérivée première

On suppose que la fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ est dérivable sur un sous-ensemble convexe $D_0 \subset \Omega$. On a alors :

i) f est convexe sur D_0 si et seulement si

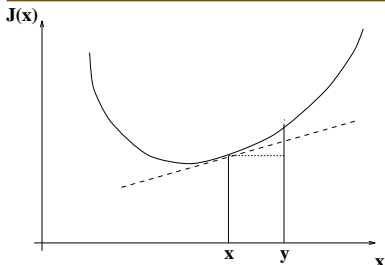
$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) .$$

ii) f est strictement convexe sur D_0 si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad \mathbf{x} \neq \mathbf{y}, \quad f(\mathbf{y}) - f(\mathbf{x}) > f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) .$$

iii) La fonctionnelle f est uniformément convexe sur D_0 si et seulement si il existe une constante $c > 0$ telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + c \|\mathbf{y} - \mathbf{x}\|_E^2 .$$



L'interprétation géométrique de

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}),$$

est que le graphe de la fonctionnelle convexe f est toujours au dessus de son plan tangent en un point quelconque du domaine D_0 .

Définition 3.2.5

Soit une fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ dérivable sur l'ouvert Ω .

L'application dérivée $f' : \Omega \subset E \rightarrow \mathcal{L}(E, \mathbb{R})$ est dite **monotone sur le sous-ensemble** $D_0 \subset \Omega$ si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \geq 0 .$$

L'application dérivée f' est dite **strictement monotone sur le sous-ensemble** $D_0 \subset \Omega$ si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \mathbf{x} \neq \mathbf{y}, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) > 0 .$$

L'application dérivée f' est dite **fortement monotone sur le sous-ensemble** $D_0 \subset \Omega$ si et seulement si il existe une constante $c > 0$ telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \geq 2c \|\mathbf{y} - \mathbf{x}\|_E^2 .$$

Proposition 3.2.6 – Relations entre convexité et monotonie de la dérivée première

On suppose que $f \in \mathcal{C}^1(\Omega)$ sur l'ouvert Ω . On a alors :

- i) La fonctionnelle f est convexe sur le sous-ensemble convexe $D_0 \subset \Omega$ si et seulement si l'application dérivée f' est monotone sur D_0 .
- ii) La fonctionnelle f est strictement convexe sur le sous-ensemble convexe $D_0 \subset \Omega$ si et seulement si l'application dérivée f' est strictement monotone sur D_0 .
- iii) La fonctionnelle f est uniformément convexe sur le sous-ensemble convexe $D_0 \subset \Omega$ si et seulement si l'application dérivée f' est fortement monotone sur D_0 (la constante $c > 0$ intervenant dans la définition de la convexité uniforme correspondant à la constante $c > 0$ introduite dans la définition de la forte monotonie de la dérivée).

Théorème 3.2.7 – Relations entre convexité et positivité de la dérivée seconde

On suppose que la fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ est deux fois dérivable dans un ouvert Ω de l'espace vectoriel normé E , et soit D_0 une partie convexe de Ω .

- i) La fonctionnelle f est convexe sur le sous-ensemble convexe $D_0 \subset \Omega$ si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) \geq 0 .$$

- ii) Si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad \mathbf{x} \neq \mathbf{y}, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) > 0 ,$$

alors la fonctionnelle f est strictement convexe sur D_0 .

- iii) La fonctionnelle f est uniformément convexe sur le sous-ensemble convexe $D_0 \subset \Omega$ si et seulement si il existe une constante $c > 0$ telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) \geq 2c \|\mathbf{y} - \mathbf{x}\|_E^2 .$$

La condition (ii) ci-dessus n'est qu'une condition suffisante, la réciproque étant inexacte.

Théorème 3.2.8 – Relations entre convexité et positivité de la dérivée seconde

On suppose que la fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ est deux fois dérivable dans un ouvert convexe Ω de l'espace vectoriel normé E .

- i) La fonctionnelle f est convexe sur le sous-ensemble convexe Ω si et seulement si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est semi-définie positive
- ii) Si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est définie positive, alors la fonctionnelle f est strictement convexe sur Ω .

Exercice 3.2.2.

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = ax_1^2 + 2bx_1x_2 + cx_2^2$. Donner une condition nécessaire et suffisante sur $(a, b, c) \in \mathbb{R}^3$ pour que f soit convexe sur \mathbb{R}^2 .

Théorème 3.2.8 – Relations entre convexité et positivité de la dérivée seconde

On suppose que la fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ est deux fois dérivable dans un ouvert convexe Ω de l'espace vectoriel normé E .

- i) La fonctionnelle f est convexe sur le sous-ensemble convexe Ω si et seulement si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est semi-définie positive
- ii) Si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est définie positive, alors la fonctionnelle f est strictement convexe sur Ω .

Exercice 3.2.2.

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = ax_1^2 + 2bx_1x_2 + cx_2^2$. Donner une condition nécessaire et suffisante sur $(a, b, c) \in \mathbb{R}^3$ pour que f soit convexe sur \mathbb{R}^2 .
- L'application $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g(x) = x_1^2 + x_1x_2 + x_2^2 + 3|x_1 + x_2 + 3|$ est-elle convexe, voire strictement convexe, sur \mathbb{R}^2 ?

Théorème 3.2.8 – Relations entre convexité et positivité de la dérivée seconde

On suppose que la fonctionnelle $f : \Omega \subset E \rightarrow \mathbb{R}$ est deux fois dérivable dans un ouvert convexe Ω de l'espace vectoriel normé E .

- i) La fonctionnelle f est convexe sur le sous-ensemble convexe Ω si et seulement si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est semi-définie positive
- ii) Si $\forall \mathbf{x} \in \Omega$, $f''(\mathbf{x})$ est définie positive, alors la fonctionnelle f est strictement convexe sur Ω .

Exercice 3.2.2.

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = ax_1^2 + 2bx_1x_2 + cx_2^2$. Donner une condition nécessaire et suffisante sur $(a, b, c) \in \mathbb{R}^3$ pour que f soit convexe sur \mathbb{R}^2 .
- L'application $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g(x) = x_1^2 + x_1x_2 + x_2^2 + 3|x_1 + x_2 + 3|$ est-elle convexe, voire strictement convexe, sur \mathbb{R}^2 ?
- L'application $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $h(x) = e^{\frac{1}{2}(x_1^2 + \dots + x_n^2)}$ est-elle convexe, voire strictement convexe, sur \mathbb{R}^n ?