Chapitre 2

Fonctions convexes

En analyse convexe, on considère très souvent des fonctions prenant des valeurs dans l'ensemble des réels auxquels on rajouter l'infini, i.e. $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Un avantage est de pouvoir inclure directement les contraintes dans la fonctionnelle optimisée, c'est-à-dire remplacer

$$\min_C f(x)$$

où K est un ensemble convexe d'un espace vectoriel E par le problème sans contraintes

$$\min_E f(x) + i_C(x), \text{ où } i_C(x) = \begin{cases} 0 \text{ si } x \in C \\ +\infty \text{ sinon} \end{cases}$$

Plus fondamentalement, des fonctions convexes prenant la valeur $+\infty$ apparaissent de manière très naturelle lorsqu'on s'intéresse à la transformée de Legendre-Fenchel, qui est unanalogue de la transformée de Fourier en analyse convexe. [1] L'ensemble $\mathbb{R} \cup \{\pm\infty\}$ est muni des règles de calcul intuitives suivantes :

$$\begin{aligned} &\forall a \in \mathbb{R}, \quad a + (+\infty) = +\infty \\ &\forall a > 0, \quad a \times (+\infty) = +\infty \\ &\forall a < 0, \quad a \times (+\infty) = -\infty \end{aligned}$$

On fera en sorte de ne jamais faire apparaître de quantités indéterminées de la forme $0 \times (+\infty)$ ou $(+\infty) - (+\infty)$).

2.1 Définition et propriétés élémentaires

Définition 7. Soit E un espace vectoriel et $f : E \to \overline{\mathbb{R}}$. On appelle

(i) *domaine de f*, noté $\mathrm{dom}(f)$, le sous-ensemble de E où f prend des valeurs finies :

$$\mathrm{dom}(f) = \{x \in E,\ f(x) \neq +\infty\};$$

1. Par exemple, on verra que la transformée de Legendre-Fenchel d'une norme est une fonction à valeurs dans $\{0, +\infty\}$.

(ii) *épigraphe de* f la partie de l'espace produit $E \times \mathbb{R}$ au-dessus du graphe de f, i.e.
$$\text{epi}(f) = \{(x,t) \in E \times \mathbb{R};\, t \geq f(x)\}.$$

Une fonction $f : E \to \overline{\mathbb{R}}$ est *propre* ssi $\text{dom}(f) \neq \emptyset$.

Définition 8. On appelle *fonction convexe* une fonction $f : E \to \overline{\mathbb{R}}$ dont l'épigraphe est un sous-ensemble convexe de $E \times \mathbb{R}$.

Remarque 2. On dit que $f : E \to \mathbb{R} \cup \{-\infty\}$ est concave si $-f$ est convexe, et tout ce qu'on va dire s'applique aussi bien au fonctions concaves qu'aux fonctions convexes, à ce changement de signe près.

Proposition 13. *Une fonction $f : E \to \overline{\mathbb{R}}$ est convexe si et seulement si* $\text{dom}(f)$ *est convexe et si pour tout x, y dans* $\text{dom}(f)$ *et tout $\alpha \in [0,1]$,*

$$f\left((1-\alpha)x + \alpha y\right) \leq (1-\alpha)f(x) + \alpha f(y) \tag{2.1}$$

Démonstration. Soit f une fonction convexe. Alors,

$$\text{dom}(f) = \{x \in \mathbb{R} \mid \exists t \in \mathbb{R}, (x,t) \in \text{epi}(f)\} = \Pi_E(\text{epi}(f))$$

où $\Pi_E : E \times \mathbb{R} \to \mathbb{R}$ est l'application affine $\Pi_E(x,t) = x$. Ainsi, $\text{dom}(f)$ est convexe comme image d'un convexe par une application affine. Ensuite, pour tout points $x, y \in E$, les points $x' := (x, f(x))$ et $y' := (y, f(y))$ appartiennent à $\text{epi}(f)$. Ainsi, par convexité de l'épigraphe, pour tout $\alpha \in [0,1]$, le point

$$z' = (1-\alpha)x' + \alpha y' = ((1-\alpha)x + \alpha y, (1-\alpha)f(x) + \alpha f(y))$$

est lui aussi dans l'épigraphe de f, ce qui se traduit par (2.1). La réciproque se démontre de la même manière. □

Cette proposition permet de déduire des opération préservant la convexité des fonctions similaire aux opérations préservant la convexité des ensembles. Nous nous contentons d'énoncer les plus importantes.

Proposition 14. *(i) Si $(f_i)_{i \in I}$ est une famille quelconque de fonctions convexes sur E, alors la fonction $x \mapsto \sup_{i \in I} f_i(x)$ est également convexe.*

(ii) Soit $A : F \to E$ une application affine, et f une fonction convexe sur E. Alors, la fonction $f \circ A$ est convexe.

(iii) Si $f_1, \ldots, f_N$ sont des fonctions convexes et $\lambda_1, \ldots, \lambda_N \geq 0$, la fonction $\sum_{i=1}^N \lambda_i f_i$ est convexe.

Exemple 6. (i) Soit C un sous-ensemble de E, et $\mathrm{i}_C : E \to \mathbb{R}$ sa fonction indicatrice, définie par $\mathrm{i}_C(x) = 0$ si $x \in C$ et $+\infty$ si $x \notin C$. Alors, i_C est convexe si et seulement si C est convexe.

(ii) Toute norme $\|\|$ sur E est convexe.
(iii) Toute forme linéaire (même discontinue) sur E est convexe.
(iv) Les sous-niveaux d'une fonction convexe sont convexes (cela se vérifie facilement manuellement, mais on peut aussi utiliser

$$\{f \leq t_0\} = \Pi_E(\mathrm{epi}(f) \cap \{(x,t) \mid t \geq t_0\})$$

où la projection $\Pi_E(x,t) = x$ est linéaire). La réciproque est fausse : il existe des fonctions non convexes dont les sous-niveaux sont tous convexes. Par exemple, les sous-niveaux de toute fonction monotone sur $\mathbb{R}$ sont convexes.
(v) Soit X une partie bornée de E. La fonction $x \mapsto \sup_{p\in X} \|x-p\|$ est convexe.
(vi) Si f est convexe et K est un ensemble convexe, la fonction $g = f + \iota_K$ est convexe ($g(x) = f(x)$ si $x \in K$, $g(x) = +\infty$ sinon). De plus, les deux problèmes d'optimisation suivants sont équivalents

$$\min_{x\in C} f(x) \Longleftrightarrow \min_{x\in E} f(x) + \iota_C(x).$$

(vii) La composition de fonctions convexes n'est pas nécessairement convexe. Par exemple $f, g : \mathbb{R} \to \mathbb{R}$, $f(x) = \exp(-x)$ et $g(x) = x^2$ sont convexes alors que $f \circ g(x) = exp(-x^2)$ n'est pas convexe.

Définition 9. Soient E un espace vectoriel normé. Une fonction convexe $f : E \mapsto \overline{\mathbb{R}}$ est *strictement convexe* si pour tout $x \neq y \in \mathrm{dom}(f)$ et tout $\lambda \in]0,1[$,

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) \tag{2.2}$$

En exercice, on pourra démontrer le lemme suivant.

Lemme 15. *Soit $f : E \to \overline{\mathbb{R}}$ atteignant son minimum $m = \min_E f$.*
(i) Si f est convexe, alors $\{x \in E \mid f(x) = m\}$ est convexe.
(ii) Si f est strictement convexe, alors $\{x \in E \mid f(x) = m\}$ est un singleton.

2.2 Continuité et lipschitzité des fonctions convexes

Pour parler de continuité et a fortiori de Lipschitzité, on supposera dans cette section que E est un espace vectoriel normé.

Définition 10. Une fonction $f : \Omega \subseteq E \to \mathbb{R}$ est localement lipschitzienne sur un ouvert Ω de E si tout point de Ω admet un voisinage sur lequel f est lipschitizienne, c'est-à-dire

$$\forall x_0 \in \Omega, \exists \delta > 0, \exists M \in \mathbb{R}, \quad \forall x, y \in \mathrm{B}(x_0, \delta), |f(x) - f(y)| \leq M \|x-y\|.$$

Proposition 16. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe sur un espace vectoriel normé et Ω un ouvert de E. Si f est bornée supérieurement sur Ω, alors elle est localement lipschitzienne sur cet ouvert.*

Remarque 3. Comme on le verra, la démonstration permet d'estimer assez précisément la constante de Lipschitz M en fonction de la borne sur f. En revanche, la constante de Lipschitz peut exploser au bord du domaine. Considérons par exemple la fonction convexe $f : x \in [0,1] \mapsto -\sqrt{x}$. Alors, $\lim_{x\to 0, x\neq 0} f'(x) = -\infty$, et f n'est pas localement lipschitzienne au voisinage de 0.

Remarque 4. On peut se servir de cette proposition pour montrer que toute forme linéaire f sur E est continue en l'origine si et seulement si elle est globalement lipschitzienne. Un des sens est évident (globalement lipschitzienne $\Longrightarrow$ continue en l'origine). Pour l'autre il faut remarquer que si f est continue en l'origine, alors elle est bornée à son voisinage et (par la proposition) L-lipschitzienne sur une boule $\mathrm{B}(0,r)$. Alors, pour $x_1, x_2 \in E$, et en posant $R = \max(\|x_1\|, \|x_2\|)$,

$$|f(x_1) - f(x_2)| = \frac{R}{r}\left|f(\frac{r}{R}x_1) - f(\frac{r}{R}x_2)\right| \leq \frac{LR}{r}\left\|\frac{r}{R}x_1 - \frac{r}{R}x_2\right\| = L\|x_1 - x_2\|.$$

Remarque 5. L'hypothèse que f est majorée est cruciale ! Par exemple, si l'espace E est de dimension infinie, on peut construire une application linéaire f non continue sur E. (Par exemple, on peut considérer $\mathbb{R}[X]$ muni de la norme ℓ^1 des coefficients, et $f : P \in \mathbb{R}[X] \mapsto P'(1)$. Alors, $P_n = \frac{1}{n}X^n$ converge vers 0, tandis que $f(P_n) = 1$. Ainsi, f est linéaire et discontinue.) Alors, $\mathrm{dom}(f) = E$, c'est-à-dire que f est finie partout, et pourtant f n'est continue en aucun point.

Avant de démontrer la proposition 16, on va démontrer un résultat intermédiaire pour une fonction convexe que l'on suppose bornée supérieurement et inférieurement.

Lemme 17. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe telle que $|f| \leq M$ sur une boule $\mathrm{B}(x_0, 2\delta)$. Alors, f est $\frac{2M}{\delta}$-lipschitzienne sur la boule $\mathrm{B}(x_0,\delta)$.*

Démonstration. Soit x, y deux points de la boule $\mathrm{B}(x_0,\delta)$. Pour comparer $f(x)$ et $f(y)$ et montrer la propriété de lipschitzité, on va prolonger le segment $[x,y]$ dans la direction de y et utiliser la borne sur f sur la boule $\mathrm{B}(x_0, 2\delta)$. Posons $\alpha = \|x-y\|$ et $z := y + \frac{\delta}{\alpha}(y-x)$. Le point z construit de cette manière appartient à la boule $\mathrm{B}(x_0,2\delta)$ car

$$\|z - x_0\| \leq \|z - y\| + \frac{\delta}{\alpha}\|y - x\| \leq 2\delta.$$

En utilisant la définition de z, on peut réécrire le point y comme combinaison convexe de x et z. La relation $(1+\delta/\alpha)y = (\delta/\alpha)x + z$ implique que

$$y = \frac{\delta/\alpha}{1+\delta/\alpha}x + \frac{1}{1+\delta/\alpha}z,$$

où la somme des deux coefficients vaut 1. Ainsi, par convexité de f,

$$\begin{aligned} f(y) &\leq \frac{\delta/\alpha}{1+\delta/\alpha}f(x) + \frac{1}{1+\delta/\alpha}f(z) \\ \text{i.e.}\quad f(y) - f(x) &\leq \frac{-1}{1+\delta/\alpha}f(x) + \frac{1}{1+\delta/\alpha}f(z) \end{aligned}$$

On peut maitenant utiliser la borne sur $|f|$:

$$f(y) - f(x) \leq \frac{2M}{1+\delta/\alpha} \leq \frac{2M}{\delta}\alpha = \frac{2M}{\delta}\|x-y\|$$

En inversant les rôles de x et y on finit de démontrer la borne sur la constante de Lipschitz de f. □

Démonstration de la Proposition 16. Soit x_0 un point de l'ouvert Ω et $\delta > 0$ tel que $\mathrm{B}(x_0, 2\delta) \subseteq \Omega$. Par hypothèse, la fonction f qu'on considère est bornée supérieurement, i.e. $f \leq M_0$ sur la boule $\mathrm{B}(x_0, 2\delta)$. Pour tout point x dans la boule $\mathrm{B}(x_0, \delta)$, le point $2x_0 - x$ est aussi dans la boule $\mathrm{B}(x_0, 2\delta)$, de sorte que

$$f(x_0) \leq \frac{1}{2}(f(x) + f(2x_0 - x)) \leq \frac{1}{2}(f(x) + M_0).$$

Ainsi, $f(x) \geq 2f(x_0) - M_0$, et la fonction est donc bornée inférieurement sur $\mathrm{B}(x_0, \delta)$. On peut donc appliquer le lemme, et en déduire que f est lipschitzienne sur la boule $\mathrm{B}(x_0, \delta/2)$. Ceci étant vrai pour tout x_0, on en déduit que f est localement Lipschitzienne sur Ω. □

Le résultat suivant montre que si f est bornée au voisinage d'un point, alors elle automatiquement continue sur l'intérieur de son domaine. La convexité permet de partir d'une hypothèse de régularité très faible (f bornée au voisinage d'un point) et d'en déduire un résultat de régularité très fort (f localement lipschitzienne sur l'intérieur de son domaine).

Proposition 18. *Soit $f : E \to \mathbb{R}$ une fonction convexe sur un espace vectoriel normé E. S'il existe un ouvert sur lequel f est borné, alors f est localement Lipschitz sur* $\mathrm{int}(\mathrm{dom}(f))$.

Démonstration. Soit $\mathrm{B}(x, \delta)$ une boule sur laquelle $|f| \leq M$. On va utiliser cette hypothèse pour démontrer que f est localement majorée dans l'intérieur de son domaine, puis conclure avec la proposition précédente. Soit y un point de $\Omega = \mathrm{int}(\mathrm{dom}(f))$. L'ensemble Ω étant ouvert, il existe $t > 0$ petit tel que le point $z := y + t(y - x)$ soit dans Ω. Par construction, le point y appartient au segment $[x, z]$. Plus précisément, comme $(1 + t)y = z + tx$ on a

$$y = \frac{t}{1+t}x + \frac{1}{1+t}z \tag{2.3}$$

$$= (1-\alpha)x + \alpha z \tag{2.4}$$

avec $\alpha = 1/(1 + t)$. Ainsi, on a

$$(1-\alpha)\mathrm{B}(x, \delta) + z = \mathrm{B}(y, (1-\alpha)\delta)$$

Montrons que $f \leq \max(M, f(z))$ sur la boule $B := \mathrm{B}(y, (1-\alpha)\delta)$. Par définition de la somme de Minkowski, pour tout point $w_y \in B$, il existe $w_x \in \mathrm{B}(x, \delta)$ tel que $w_y = (1-\alpha)w_x + \alpha z$. D'où

$$f(w_y) \leq (1-\alpha)f(w_x) + \alpha f(z) \leq \max(M, f(z)).$$

La fonction est donc bornée supérieurement au voisinage de y. Par la proposition précédente, f est donc lipschitzienne au voisinage de y. Ceci étant vrai pour tout $y \in \Omega$, on en conclut que f est localement lipschtzienne sur Ω. $\square$

Corollaire 19. *Soit $f : E \to \mathbb{R}$ une fonction convexe sur un espace vectoriel normé E. S'il existe un point $x_0 \in \mathrm{dom}(f)$ où f est continue, alors f est localement Lipschitz sur $\mathrm{int}(\mathrm{dom}(f))$.*

2.3 Dérivées directionnelles

Dans cette section, on utilise les propriétés "algébriques", c'est-à-dire sans topologie, des dérivées directionnelles d'une fonction convexe. L'espace E est donc un espace vectoriel quelconque, qui n'est pas nécessairement normé. Comme on utilise uniquement la structure linéaire de l'espace, on ne doit pas s'attendre à pouvoir en déduire des informations sur la régularité (même la continuité !) des fonctions considérées.

Définition 11. Soit E un espace vectoriel, $f : E \to \mathbb{R}$ une fonction, x un point du domaine de f et $v \in E$ une direction. On pose, si la limite existe,

$$f^+(x; v) = \lim_{\varepsilon \to 0^+} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \tag{2.5}$$

Proposition 20. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe, et $x \in \mathrm{dom}(f)$. Alors, la dérivée directionnelle $v \in E \mapsto f^+(x; v) \in \mathbb{R} \cup \{\pm\infty\}$ est bien définie et de plus,*

$$f^+(x; v) = \inf_{\varepsilon > 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \tag{2.6}$$

Remarque 6. La limite définissant $f^+(x; \pm)$ peut prendre les valeurs $\pm\infty$.

(i) Le fait de pouvoir remplacer la limite (2.5) par un infimum (2.6) implique que $f^+(x; v) = +\infty$ si et seulement si la demi-droite $\{x + tv \mid t > 0\}$ n'intersecte pas le domaine de f. Par contraposée, si x est dans l'intérieur du domaine de f (ou dans l'intérieur algébrique du domaine de f, cf TD), alors $f^+(x; v) < +\infty$.

(ii) Il est facile de construire des exemples de fonctions tels que $f^+(x; v) = -\infty$. Par exemple $f(x) = -\sqrt{x}$, $x = 0$ et $v = 1$. Alors $f^+(x; v) = \lim_{t \to 0} -f'(t) = -\infty$.

Lemme 21. *La fonction $\varepsilon \mapsto \frac{1}{\varepsilon}(f(x + \varepsilon v) - f(x))$ est croissante.*

Démonstration. Soient $\varepsilon_2 \geq \varepsilon_1 \geq 0$, et supposons de plus que le point $x + \varepsilon_2 v$ appartient à $\mathrm{dom}(f)$ (sinon $f(x + \varepsilon_2 v) = +\infty$ et il n'y a rien à démontrer). Par hypothèse, le point $x + \varepsilon_1$ appartient au segment $[x, x + \varepsilon_2]$ et donc au domaine de f par convexité de celui-ci. Plus précisément on a :

$$x + \varepsilon_1 v = (1 - \varepsilon_1/\varepsilon_2)x + \varepsilon_1/\varepsilon_2(x + \varepsilon_2 v),$$

de sorte qu'en utilisant la convexité de f on obtient

$$f(x + \varepsilon_1 v) \leq (1 - \varepsilon_1/\varepsilon_2)f(x) + \varepsilon_1/\varepsilon_2 f(x + \varepsilon_2 v).$$

Ainsi,

$$\frac{f(x + \varepsilon_1 v) - f(x)}{\varepsilon_1} \leq \frac{f(x + \varepsilon_2 v) - f(x)}{\varepsilon_2}.$$ □

Démonstration de la proposition 20. Le ratio $\frac{1}{\varepsilon}(f(x + \varepsilon v) - f(x))$ étant décroissant lorsque $\varepsilon \to 0^+$, il admet une limite dans $\mathbb{R} \cup \{\pm\infty\}$ donnée par (2.6). Si $f^+(x; v) < +\infty$, alors il existe ε tel que $f(x + \varepsilon d) < +\infty$, auquel cas f est finie sur $[x, x + \varepsilon]$ par convexité. □

Proposition 22. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe,et $x \in \mathrm{dom}(f)$. Alors,*

(i) La fonction $g : v \mapsto f^+(x, v)$ est positivement 1-homogène ($g(\lambda v) = \lambda g(v)$ pour $\lambda > 0$) et sous-linéaire ($g(v + w) \leq g(v) + g(w)$).

(ii) Propriété de monotonie : pour tout $x, y \in \mathrm{dom}(f)$,

$$f^+(x; y - x) \leq f^+(y; y - x) \tag{2.7}$$

Démonstration. (i) Il est facile de voir que $g = f^+(x, .)$ est 1-homogène. Soient u, v dans E, et montrons que $g(u + v) \leq g(u) + g(v)$. On peut supposer que que pour $\varepsilon > 0$ assez petit, $x + \varepsilon u$ et $x + \varepsilon v$ appartiennent à $\mathrm{dom}(f)$, sinon il n'y a rien à montrer. Alors,

$$x + \varepsilon(u + v) = \frac{x + 2\varepsilon u}{2} + \frac{x + 2\varepsilon v}{2}$$

et, par convexité,

$$\frac{1}{\varepsilon}(f(x + \varepsilon(u + v)) - f(x)) \leq \frac{1}{2\varepsilon}(f(x + 2\varepsilon u) - f(x)) + \frac{1}{2\varepsilon}(f(x + 2\varepsilon v) - f(x)).$$

En passant à la limite on obient l'inégalité voulue.
(ii) Cette propriété correspond à la croissance des pentes pour les fonctions convexes sur un segment de $\mathbb{R}$. □

Proposition 23. *Soit $f : E \to \overline{\mathbb{R}}$ convexe, et $x \in \mathrm{dom}(f)$. De plus, la fonction $v \in E \mapsto f^+(x; v)$ est linéaire sur E si et seulement si*

$$\forall v \in E, f^+(x; v) < +\infty$$
$$\forall v \in E, f^+(x; v) = -f^+(x; -v).$$

Remarque 7. La fonction $v \mapsto f^+(x;v)$ peut tout à fait être linéaire et discontinue ! En effet, pour toute forme linéaire f sur E, on a $f^+(x;v) = f(v)$. Il suffit donc de choisir f est linéaire *non continue* en dimension infinie pour obtenir un exemple de dérivée directionnelle $v \mapsto f^+(x;v)$ qui est également linéaire et non continue.

Exemple 7. Soit $f(x) = |x|$ sur $\mathbb{R}$, alors $f^+(0;1) = 1$ et $f^+(0;-1) = 1 \neq -f^+(0,1)$.

Ce corollaire se déduit de la sous-linéarité de $f^+(x;\cdot)$ et des deux lemmes suivants.

Lemme 24. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe, et $x \in \mathrm{dom}(f)$. Alors, $f^+(x,v) \geq -f^+(x,-v)$.*

Démonstration. On peut supposer $f^+(x,v) < +\infty$ ou $f^+(x,-v) < +\infty$ car sinon l'inégalité est triviale. Alors, $x - \varepsilon v$ et $x + \varepsilon v$ appartiennent à $\mathrm{dom}(f)$ pour $\varepsilon > 0$ assez petit. Par la convexité de f,

$$f(x) = f\left(\frac{x+\varepsilon v}{2} + \frac{x - \varepsilon v}{2}\right) \leq \frac{1}{2} f(x+\varepsilon v) + \frac{1}{2} f(x - \varepsilon v)$$

Ainsi,

$$\frac{f(x+\varepsilon v) - f(x)}{\varepsilon} \geq -\frac{f(x-\varepsilon v) - f(x)}{\varepsilon},$$

ce qui donne résultat par passage à la limite. $\square$

Lemme 25. *Une fonction sous-linéaire $g : E \to \overline{\mathbb{R}}$ est linéaire si et seulement si elle est partout finie et si $g(v) = -g(-v)$ pour tout v dans E.*

Démonstration. Par sous-linéarité de g,

$$\begin{aligned} &g(v+w) \leq g(v) + g(w) \\ &g(-(v+w)) \leq g(-v) + g(-w) = -g(v) - g(w), \end{aligned}$$

où l'on a utilisé l'hypothèse pour la dernière égalité. Ainsi,

$$g(v) + g(w) \leq -g(-(v+w)) = g(v+w) \leq g(v) + g(w),$$

et toutes les inégalités doivent donc être des égalités. En particulier, on obtient l'additivité de g : $g(v+w) = g(v) + g(w)$. Comme on sait de plus que $g(\lambda v) = \lambda g(v)$, g est linéaire. $\square$

On conclut cette partie par une caractérisation de la convexité utilisant uniquement la notion de dérivée directionnelle.

Proposition 26. *Soit $f : X \subseteq E \to \mathbb{R}$ une fonction sur un ouvert convexe X. On suppose que pour tout $x \in X$, l'application dérivée directionnelle*

$$\mathrm{D}_x f : v \in E \mapsto \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f(x+\varepsilon v) - v)$$

est bien définie et linéaire. Alors, les propriétés suivantes sont équivalentes :

(i) f *est convexe ;*
(ii) *pour tout* $x, y \in X$, $f(y) \geq f(x) + \mathrm{D}_x f(y-x)$ *;*
(iii) *pour tout* $x, y \in X$, $(\mathrm{D}_x f - \mathrm{D}_y f)(x-y) \geq 0$.

Démonstration. (i) $\Longrightarrow$ (ii) : Comme la fonction f est convexe,

$$\begin{aligned} D_x f(y-x) &= \lim_{t \to 0^+} \frac{1}{t}(f(x+t(y-x)) - f(x)) \\ &= \lim_{t \to 0^+} \frac{1}{t}(f((1-t)x+ty) - f(x)) \\ &\leq \lim_{t \to 0^+} \frac{1}{t}((1-t)f(x)+tf(y) - f(x)) = f(y) - f(x) \end{aligned}$$

(ii) $\Longrightarrow$ (iii) : Il suffit de sommer les inégalités

$$\begin{aligned} f(y) &\geq f(x) + \mathrm{D}_x f \cdot (y-x) \\ f(x) &\geq f(y) + \mathrm{D}_y f \cdot (x-y) \end{aligned}$$

(iii) $\Longrightarrow$ (i) : Soient x, y dans X et $\phi(\lambda) = f(x_\lambda) - (1-\lambda)f(x) - \lambda f(y)$, où on a posé $x_\lambda := (1-\lambda)x + \lambda y$. La fonction ϕ est différentiable en tout $\lambda \in [0,1]$ et $\phi(0) = \phi(1) = 0$. Montrer que f est convexe sur le segment $[x, y]$ revient à montrer que $\phi(\lambda) \leq 0$ sur $[0,1]$. Supposons le contraire, et considérons λ_0 un point du segment ouvert $]0,1[$ où ϕ atteint son maximum, de sorte que $\phi'(\lambda_0) = 0$ et $\phi(\lambda_0) > 0$. Pour tout $\lambda \in [0,1]$, on a :

$$\phi'(\lambda) - \phi'(\lambda_0) = (\mathrm{D}_{x_\lambda} f - \mathrm{D}_{x_{\lambda_0}} f) \cdot (y-x)$$

De plus, si $\lambda \neq \lambda_0$,

$$y - x = \frac{1}{\lambda - \lambda_0}[((1-\lambda)x + \lambda y) - ((1-\lambda_0)x + \lambda_0 y)] = \frac{1}{\lambda - \lambda_0}(x_\lambda - x_{\lambda_0})$$

Ainsi, en utilisant l'hypothèse (iii) on obtient que si $\lambda > \lambda_0$, alors $\phi'(\lambda) \geq \phi'(\lambda_0) = 0$. La fonction ϕ devrait donc être croissante croissante sur l'intervalle $[\lambda_0, 1[$, et en particulier $\phi(1) \geq \phi(\lambda_0)$. Ceci contredit l'inégalité $\phi(\lambda_0) > 0 = \phi(1)$. Par l'absurde, on en déduit que $\phi \leq 0$, puis que f est convexe sur le segment $[x, y]$ et enfin qu'elle est convexe sur l'ouvert X. □

2.4 Différentiabilité au sens de Gâteaux et de Fréchet

Définition 12. Soient E un espaces vectoriel normé. Une fonction $f : E \to \overline{\mathbb{R}}$ est dite Gâteaux-différentiable en $x \in \mathrm{dom}(f)$ si elle admet une dérivée directionnelle en x dans toutes les directions $v \in E$

$$\mathrm{D}_x f(v) := \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(x+tv) \in \mathbb{R},$$

et si l'application $v \mapsto \mathrm{D}_x f(v)$ est linéaire continue sur E.

Remarque 8. La différentiabilité au sens de Gâteaux est une notion assez faible. Par exemple, considérons la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x_1, x_2) = \begin{cases} 1 \text{ si } x_1 \neq 0 \text{ et } x_2 = x_1^2 \\ 0 \text{ sinon} \end{cases}$$

Les dérivées directionnelles de f en $(0, 0)$ sont toutes nulles, de sorte que f est Gâteaux-différentiable en ce point avec $\mathrm{D}_{(0,0)} f = 0$. Cependant, la fonction f n'est même pas continue en $(0, 0)$!

Proposition 27. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe continue en $x \in E$. Alors,*

$$\begin{aligned} f \text{ est Gâteaux-différentiable en } x &\iff v \in E \mapsto f^+(x; v) \text{ est linéaire} \\ &\iff \forall v \in E, \quad f^+(x; v) = -f^+(x; -v) \end{aligned}$$

Dans ce cas, on a $\mathrm{D}_x f = f^+(x; \cdot)$.

On sait déjà que l'application $v \mapsto f^+(x; v)$ est linéaire sous la deuxième hypothèse, il suffit donc d'appliquer le lemme suivant.

Lemme 28. *Soit $f : E \to \overline{\mathbb{R}}$ une fonction convexe continue en $x \in E$. Alors l'application $v \in E \mapsto f^+(x, v)$ est continue.*

Démonstration. Comme f est continue en x, elle est localement bornée au voisinage de x et donc M-Lipschitz dans un voisinage de x. En particulier,

$$f(x_0 + \varepsilon v) - f(x_0) \leq M\varepsilon \|v\|,$$

d'où $g(v) := f^+(x; v) \leq M \|v\|$. Par sous-additivité de la fonction g, pour $v, h \in E$,

$$g(v) - g(h) \leq g(v + h) \leq g(v) + g(h),$$

on en déduit la continuité de g : $g(v) - M \|h\| \leq g(v + h) \leq g(v) + M \|h\|$. □

Définition 13. Une fonction $f : E \to \overline{\mathbb{R}}$ est dite *Fréchet-différentiable* en $x \in \mathrm{dom}(f)$ si elle est Gâteaux-différentiable en x et si

$$\lim_{v \to 0, v \neq 0} \frac{|f(x + v) - f(x) - \mathrm{D}_x f(v)|}{\|v\|} = 0 \tag{2.8}$$

ou de manière plus compacte, $f(x + v) = f(x) + \mathrm{D}_x f(v) + \mathrm{o}(\|v\|)$.

Remarque 9. La différentiabilité au sens de Fréchet est la notion habituelle de différentiabilité. Les implications suivantes sont vraies (et immédiates) : Fréchet différentiabilité $\Longrightarrow$ Gâteaux-différentiabilité $\Longrightarrow$ linéarité de l'application $v \mapsto f^+(x; v)$. En revanche, les implications réciproques sont fausses sans hypothèses supplémentaires.

Remarque 10. La Fréchet-différentiabilité implique évidemment la continuité. Ainsi, la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ considérée dans l'exemple précédent, qui est Gâteaux-différentiable en $(0, 0)$ mais discontinue en ce point, n'est pas Fréchet-différentiable.

2.5 Théorèmes de différentiabilité presque partout

Motivation. Soit H un espace de Hilbert. Étant donné un compact convexe $K \subseteq E$ et $x^* \in E$, on s'intéresse au problème de programmation linéaire suivant :

$$\max_{x \in K} \langle x^* | x \rangle \tag{2.9}$$

On note $f : x^* \in E \to \mathbb{R}$ la fonction valeur du problème de programmation linéaire, qui est convexe comme maximum de fonctions linéaires. De plus, si x est une solution du problème (2.9), c'est-à-dire $x \in K$ et $f(x^*) = \langle x^* | x \rangle$, on a

$$\begin{aligned} f^+(x^*, v^*) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f^+(x^* + \varepsilon v^*) - f(x^*)) \\ &\geq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(\langle x^* + v^* | x \rangle - \langle x^* | x \rangle) \geq \langle v^* | x \rangle \end{aligned}$$

On s'intéresse maintenant à l'unicité du maximiseur de (2.9). Supposons qu'il existe $x \neq y \in K$ tel que $f(x^*) = \langle x^* | x \rangle = \langle x^* | y \rangle$. Alors, par le raisonnement précédent,

$$f^+(x^*, v^*) \geq \max(\langle v^* | x \rangle, \langle v^* | y \rangle).$$

L'application $v^* \in E^* \mapsto f^+(x^*; v^*)$ ne peut alors pas être linéaire, et f n'est donc pas différentiable en x. Étudier la différentiabilité de f en x^* nous apprend donc des choses sur l'unicité de la solution au problème de programmation linéaire (2.9).

2.5.1 Différentiabilité des fonctions convexes sur $\mathbb{R}$

Théorème 29. *Soit $f : \mathbb{R} \to \overline{\mathbb{R}}$ une fonction convexe, et $I = \mathrm{int}(\mathrm{dom}(f))$. Alors, l'ensemble des points de I où f n'est pas dérivable est au plus dénombrable.*

Démonstration. Soit $f'_d(x) = f^+(x; 1)$ et $f'_g(x) = -f^+(x, -1)$ les dérivées à droite et gauche. En utilisant la croissance des pentes d'une fonction convexe sur $\mathbb{R}$, on peut montrer que ces fonctions sont croissantes (exercice). Pour tout $x \leq x_0$ dans I, on a

$$f'_d(x) = \inf_{y > x} \frac{f(y) - f(x)}{y - x} \leq \frac{f(x_0) - f(x)}{x_0 - x} \leq f'_g(x_0),$$

ce qui implique l'inégalité

$$\lim_{x \to x_0^-} f'_d(x) \leq f'_g(x_0) \leq f'_d(x_0)$$

La fonction f est dérivable en x_0 si et seulement si $f'_g(x_0) = f'_d(x_0)$. Ainsi, si f n'est pas différentiable en x_0, la fonction f'_d a un saut en x_0 :

$$\lim_{x \to x_0^-} f'_d(x) < f'_d(x_0).$$

On conclut en utilisant le fait qu'une fonction croissante ne peut avoir qu'un nombre dénombrable de sauts. □

Remarque 11. Ce théorème est faux en dimension plus grande. Considérer la fonction convexe f sur $\mathbb{R}^2$ définie par $f(x_1, x_2) = |x_1|$: cette fonction n'est pas différentiable sur la droite $\{0\} \times \mathbb{R}$, qui est indénombrable.

2.5.2 Gâteaux-différentiabilité des fonctions convexes sur un espace de Banach séparable

On rappelle qu'un espace de Banach E est dit séparable si il contient un ensemble dénombrable dense.

Théorème 30 (Mazur)**.** *Soit E un espace de Banach séparable, $\Omega \subseteq E$ un ouvert convexe et $f : \Omega \to \mathbb{R}$ une fonction convexe continue. Alors, f est Gâteaux-différentiable sur un sous-ensemble dense de Ω.*

Pour démontrer ce théorème on considère une une suite $(v_n)_{n\geq 0}$ dense dans E et on introduit les ensembles

$$A_{m,n} = \{x \in \Omega \mid f^+(x, v_n) + f^+(x, -v_n) \geq 1/m\}, \qquad A = \bigcup_{m,n\geq 1} A_{m,n} \tag{2.10}$$

Le plan de la démonstration est le suivant : (a) f est Gâteaux-différentiable sur l'ensemble $\Omega \setminus A$ (b) que chacun des $A_{m,n}$ est fermé et (c) que chacun des $A_{m,n}$ est d'intérieur vide. Par théorème de Baire appliqué à un ouvert dans un espace complet (cf [**?**, p. 83]), on sait alors que A est d'intérieur vide (ou de manière équivalente que $\Omega \setminus A$ est dense).

Proposition 31. *La fonction f est Gâteaux-différentiable sur l'ensemble $\Omega \setminus A$.*

Démonstration. Comme f est continue sur Ω, par la proposition 27,

$$\begin{aligned} f \text{ n'est pas G.-différentiable en } x \in \Omega &\Longrightarrow \exists v \in E, f^+(x, v) + f^+(x, -v) > 0 \\ &\Longrightarrow \exists v \in E, \exists m > 1, f^+(x, v) + f^+(x, -v) > 2/m \\ &\Longrightarrow \exists m, n \geq 1, f^+(x, v_n) + f^+(x, -v_n) > 1/m \\ &\Longrightarrow x \in A, \end{aligned}$$

où l'on a utilisé la continuité de l'application $v \mapsto f^+(x, v)$. □

Proposition 32. *L'ensemble $A_{m,n}$ défini par* (2.10) *est fermé dans E.*

Cette proposition est une conséquence immédiate du lemme suivant donnant la semicontinuité supérieure de $x \mapsto g^+(x, v)$.

Lemme 33. *Soit $g : E \to \overline{\mathbb{R}}$ une fonction convexe continue en un point x de E, et soit (x_k) une suite qui converge vers x. Alors, $g^+(x, v) \geq \limsup_{k\to\infty} g^+(x_k, v)$.*

Démonstration. Comme g est continue en x, elle est L-lipschitzienne dans un voisinage de x. Sans perte de généralité, on suppose que la suite (x_k) reste dans ce voisinage. Soit $\varepsilon > 0$. En utilisant la lipschitzité de g et $x_k \to x$, on a

$$\begin{aligned}\frac{1}{\varepsilon}(g(x+\varepsilon v) - f(x)) &= \frac{1}{\varepsilon}(g(x_k + \varepsilon v) - g(x_k) - 2L\,\|x - x_k\|) \\ &\geq g^+(x_k, v) - \frac{2L\,\|x - x_k\|}{\varepsilon} \\ &\geq \limsup_{k\to\infty} g^+(x_k, v).\end{aligned}$$

On en déduit le lemme en passant à l'infimum à gauche. □

Proposition 34. *L'ensemble $A_{m,n}$ défini par* (2.10) *est d'intérieur vide.*

Démonstration. On raisonne par l'absurde, et l'on suppose que l'intérieur de $A_{m,n}$ contient un point x. Alors, il existe $r > 0$ tel que $\mathrm{B}(x, r) \subseteq A_{m,n}$. Soit $x_t := x + tv_n$ et $g : t \in [0, r] \mapsto f(x_t)$. Alors,

$$\begin{aligned}&\forall t \in [0, r], \quad -f^+(x_t, -v_n) + 1/m \leq f^+(x_t, v_n) \\ &\implies \forall t \in [0, r], \quad g \text{ n'est pas différentiable en } t\end{aligned}$$

Ceci contredit le théorème (29), qui affirme que l'ensemble de non-différentiabilité de g est au plus dénombrable. □

2.5.3 Fréchet-différentiabilité presque partout des fonctions convexes en dimension finie

Le comportement des fonctions convexes en dimension finie est beaucoup plus simple qu'en dimension infinie. Soit $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ une fonction convexe et x est un point de continuité de f. On va montrer la chaine d'implication suivante :

$$\begin{aligned}&f \text{ admet des dérivées partielles } \left(\tfrac{\partial f}{\partial e_i}(x)\right)_{1\leq i\leq d} \\ &\implies \text{l'application } v \mapsto f^+(x; v) \text{ est linéaire} \\ &\implies f \text{ est Gâteaux-différentiable en } x \\ &\implies f \text{ est Fréchet-différentiable en } x\end{aligned}$$

On en déduira le théorème principal de ce chapitre, affirmant qu'une fonction convexe $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ est différentiable en presque tout point de son domaine.

Remarque 12. Dans la suite, on fera souvent l'hypothèse que le domaine des fonctions considérées est d'intérieur non vide. Pour traiter le cas général, il suffit de considérer la restriction de f à l'enveloppe affine de $\mathrm{dom}(f)$.

Proposition 35. *Soit E un espace de dimension finie et $f : E \to \overline{\mathbb{R}}$ une fonction convexe. La restriction de f à l'intérieur relatif de* $\mathrm{dom}(f)$ *est localement Lipschitz.*

Démonstration de la proposition 35. On suppose également que $E = \mathbb{R}^n$. Quitte à restreindre f à l'enveloppe affine de son domaine, on suppose que $\Omega = \mathrm{int}(\mathrm{dom}(f))$ est non vide, et quitte à translater, on suppose que Ω contient l'origine. Il existe donc $r > 0$ tel que Ω contienne le cube $[-r, r]^n$. Soit $X = \{(\pm r, \dots, \pm r)\}$, de sorte que $[-r, r]^n = \mathrm{conv}(X)$ (exercice). On pose

$$M = \max_{\varepsilon_i \in \{\pm 1\}^n} f(\varepsilon_1 r, \dots, \varepsilon_n r).$$

Soit $x \in [-r, r]^n$. Comme $[-r, r]^n = \mathrm{conv}(X)$, il existe $k \geq 0$, $x_1, \dots, x_k \in X$, $\lambda_1, \dots, \lambda_k \geq 0$ de sorte que $\sum_i \lambda_i = 1$ et $x = \sum_i \lambda_i x_i$. Alors,

$$f(x) \leq \sum_{i=1}^k \lambda_i f(x_i) \leq \sum_{i=1}^k \lambda_i M = M.$$

Ainsi, f est localement bornée en un point, et par proposition 18 elle est localement lipschitzienne sur l'intérieur de son domaine. □

Proposition 36. *Soit E un espace vectoriel de dimension finie et $f : E \to \overline{\mathbb{R}}$ une fonction convexe. Si f est Gâteaux-différentiable en un point x de* $\mathrm{int}(\mathrm{dom}(f))$*, alors f est Fréchet-différentiable en ce point.*

Cette proposition est en fait une conséquence du lemme suivant, et du fait que f est localement lipschitzienne au voisinage de x.

Lemme 37. *Soit $f : \mathrm{B}(x, r) \to \mathbb{R}$, $\dim(E) < +\infty$, une fonction M-Lipschitz. Si f est Gâteaux-différentiable en x, alors elle est également Fréchet-différentiable en x.*

Démonstration. Soit $\varepsilon > 0$. Par compacité de la sphère unité S de E, il existe une famille de vecteurs $(v_i)_{1 \leq i \leq N}$ de S telle que $S \subseteq \cup_i \mathrm{B}(v_i, \varepsilon)$. Par Gâteaux-différentiabilité de f en x, pour tout $\varepsilon > 0$ et tout i, il existe δ_i tel que

$$\forall t \in [-\delta_i, \delta_i], \ \|f(x + tv_i) - (f(x) + t\mathrm{D}_x f(v_i))\| \leq \varepsilon \, |t|$$

Soit $\delta := \min_i \delta_i > 0$. Par construction des (v_i), pour tout vecteur v de S, il existe i tel que $\|v_i - v\| \leq \varepsilon$. Alors, en utilisant le caractère Lipschitz de f et de $\mathrm{D}_x f$,

$$\begin{aligned} \|f(x + tv_i) - f(x + tv)\| &\leq M \, |t| \, \varepsilon \\ \|\mathrm{D}_x f(v_i) - \mathrm{D}_x f(x + tv)\| &\leq M \, |t| \, \varepsilon \end{aligned}$$

Ainsi, pour tout $v \in S$ et $t \leq \delta$,

$$\begin{aligned} \|f(x + tv) - (f(x) + t\mathrm{D}_x f(v))\| &\leq \|f(x + tv_i) - (f(x) + t\mathrm{D}_x f(v_i))\| + 2M\varepsilon \, |t| \\ &\leq (2M + 1)\varepsilon \, |t| \end{aligned}$$

De manière équivalente, pour tout $v \in E$, $\|v\| \leq \delta$,

$$\|f(x+v) - (f(x) + \mathrm{D}_x f(v))\| \leq (2M+1)\varepsilon \|v\|,$$

et la fonction f est donc bien Fréchet-différentiable en x. $\square$

Proposition 38. *Soit E un espace vectoriel de dimension finie, et (e_i) une base de E et $f : E \to \mathbb{R}$ une fonction convexe. Alors f est Gâteaux-différentiable en $x \in \mathrm{int}(\mathrm{dom}(f))$ si et seulement si elle admet des dérivées partielles en x :*

$$\frac{\partial f}{\partial e_i}(x) = \lim_{t \to 0} \frac{f(x+te_i) - f(x)}{t}.$$

Lemme 39. *Soit $g : E \to \mathbb{R}$ une fonction sous-linéaire. Alors, l'ensemble*

$$V = \{v \in E \mid f^+(x; v) = -f^+(x; -v)\}$$

est un sous-espace vectoriel de E.

Démonstration. Par sous-linéarité, $0 = p(u + (-u)) \leq p(u) + p(-u)$, de sorte que $-p(-u) \leq p(u)$. Par construction, l'ensemble V est stable par multiplication par un scalaire. Soient $v, w \in V$. On a

$$g(v+w) \leq g(v) + g(w) = -g(-v) + -g(-w) \leq -g(-v-w) \leq g(v+w).$$

Ainsi, $v + w \in V$ et V est bien un sous-espace vectoriel de E. $\square$

Démonstration de la proposition 38. La fonction f est localement Lipschitz au voisinage de x, et $g = f^+(x; \cdot)$ est sous-linéaire. Soit $V := \{v \in E \mid g(v) = -g(-v)\}$. Par le lemme précédent, V est un sous-espace vectoriel de E, et par hypothèse $e_i \in V$ pour tout i. Ainsi, $V = E$ et $f^+(x; \cdot)$ est linéaire. $\square$

Théorème 40. *Soit E un espace vectoriel de dimension finie et $f : E \to \overline{\mathbb{R}}$ une fonction convexe. Alors, f est Fréchet-différentiable en presque tout point de* $\mathrm{int}(\mathrm{dom}(f))$.

Démonstration. Soit A l'ensemble des points de $\Omega := \mathrm{int}(\mathrm{dom}(f))$ où la fonction f n'est pas Fréchet-différentiable. Par la proposition 38, l'ensemble A est contenu dans l'intersection des ensembles

$$A_i := \left\{ x \in \Omega \mid \frac{\partial f}{\partial e_i} \text{ n'existe pas en } x \right\}.$$

Ainsi, pour montrer que A a une mesure nulle, il suffit de démontrer que chacun des A_i a une mesure nulle. Sans perte de généralité, on suppose que $E = \mathbb{R}^n$ et $i = n$, et on considère ϕ la fonction indicatrice de A_n. Par thèorème de Tonelli,

$$\lambda(A_n) = \int_{\mathbb{R}^n} \phi(x) \mathrm{d}x = \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \phi(y, x_n) \mathrm{d}x_n \mathrm{d}y$$

Or, pour tout $y \in \mathbb{R}^{n-1}$, $t \mapsto \phi(y,t)$ est la fonction indicatrice du lieu B_y de non-différentiabilité de la fonction convexe $t \in \mathbb{R} \mapsto f(y,t)$. Par le theoreme 29, B_y est dénombrable et donc de mesure de Lebesgue nulle. $\square$