Module A1

Éléments d'analyse convexe

Dans ce premier module, on rappelle quelques éléments d'analyse convexe qui nous seront utiles en optimisation convexe.

Sauf mention contraire, $\mathcal{X}$ est un espace vectoriel réel de dimension finie, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\| \cdot \|$ la norme euclidienne associée.

1 Ensembles convexes

1.1 Définition et exemples

La définition des ensembles convexes repose sur la définition suivante :

Définition 1 (Combinaison convexe I)

Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. L'élément

$$\lambda\, x_1 + (1-\lambda)\, x_2 \in \mathcal{X}$$

est appelé *combinaison convexe* de x_1 et x_2. On note $[\,x_1\,;x_2\,]$ l'ensemble des combinaisons convexes de x_1 et x_2.

Dans $\mathbb{R}$, la définition de $[\,x_1\,;x_2\,]$ coïncide avec celle de l'intervalle fermé et borné $[\,x_1\,;x_2\,]$. Dans le plan $\mathbb{R}^2$, elle coïncide avec celle d'un segment usuel, à savoir la portion de droite reliant les points x_1 et x_2.

Notons que la somme des coefficients qui interviennent dans la combinaison convexe de deux points vaut $\lambda + (1-\lambda) = 1$. Aussi, on peut généraliser cette notion à un nombre de points supérieur à deux :

Définition 2 (Combinaison convexe II)

Soit $\{x_i\}_{1 \le i \le n} \in \mathcal{X}^n$ et $\{\lambda_i\}_{1 \le i \le n} \in [\,0\,;1\,]^n$. Si

$$\lambda_1 + \cdots + \lambda_n = 1$$

alors l'élément $$\lambda_1\, x_1 + \cdots + \lambda_n\, x_n \in \mathcal{X}$$

est appelé *combinaison convexe* des x_i.

On peut maintenant donner la définition suivante :

Définition 3 (Ensemble convexe)

Soit $\mathcal{C} \subset \mathcal{X}$. On dit que $\mathcal{C}$ est *convexe* si

$$\forall\, (x_1, x_2) \in \mathcal{C}^2, \qquad \lambda\, x_1 + (1-\lambda)\, x_2 \in \mathcal{C}$$

Autrement dit, un ensemble convexe contient toutes les combinaisons convexes des paires de ses points. Il est aisé de vérifier qu'il contient aussi toutes les combinaisons convexes de tous n-uplets de ses points. On peut donc simplement dire que $\mathcal{C}$ est *stable par combinaison convexe.*

Exercice

Convexes dans $\mathbb{R}$. Montrer que les ensembles convexes de $\mathbb{R}$ sont les singletons et les intervalles (ouverts, fermés, semi-ouverts, bornés ou non).

Exemple

Quelques exemples en dimension finie. Les ensembles suivants sont convexes :

- les (sous-)espaces vectoriels ;
- les singletons ;
- les demi-espaces d'un espace vectoriel ;
- les boules (fermées ou ouvertes).

1.2 Propriétés

On va maintenant s'intéresser aux opérations préservant la convexité :

Proposition 1 (Somme)

Soit $\mathcal{C}_1$ et $\mathcal{C}_2$ deux ensembles convexes de $\mathcal{X}$. Alors

$$\mathcal{C}_1 + \mathcal{C}_2 = \big\{ x_1 + x_2 \in U \mid x_1 \in \mathcal{C}_1 \text{ et } x_2 \in \mathcal{C}_2 \big\}$$

est un ensemble convexe de $\mathcal{X}$.

Démonstration : Soit z et z' deux points de $\mathcal{C}_1 + \mathcal{C}_2$. Par définition de la somme de deux ensembles, il existe donc $(x_1, x_1') \in \mathcal{C}_1^2$ et $(x_2, x_2') \in \mathcal{C}_2^2$ tels que

$$z = x_1 + x_2 \qquad \text{et} \qquad z' = x_1' + x_2'$$

Soit $\lambda \in [\,0\,;1\,]$. Par convexité des ensembles $\mathcal{C}_1$ et $\mathcal{C}_2$, on a $\lambda\, x_1 + (1-\lambda)\, x_1' \in \mathcal{C}_1$ et $\lambda\, x_2 + (1-\lambda)\, x_2' \in \mathcal{C}_2$. Par conséquent, la somme de ces deux points appartient à $\mathcal{C}_1 + \mathcal{C}_2$. Or, puisque

$$(\lambda\, x_1 + (1-\lambda)\, x_1') + (\lambda\, x_2 + (1-\lambda)\, x_2') = \lambda\, (x_1 + x_2) + (1-\lambda)\, (x_1' + x_2')$$

on en déduit que $\lambda\, (x_1 + x_2) + (1-\lambda)\, (x_1' + x_2') \in \mathcal{C}_1 + \mathcal{C}_2$. ■

Proposition 2 (Intersection)

Soit $\mathcal{I} \subset \mathbb{R}$ et $\mathcal{C}_i$ un ensemble convexe de $\mathcal{X}$ pour tout $i \in \mathcal{I}$. Alors

$$\bigcap_{i \in \mathcal{I}} \mathcal{C}_i = \big\{ x \in U \mid \forall\, i \in \mathcal{I}, \quad x \in \mathcal{C}_i \big\}$$

est un ensemble convexe de $\mathcal{X}$.

Démonstration : Laissé au lecteur.

Remarque : En revanche, l'union de deux ensembles convexes n'est pas toujours convexe (il suffit de considérer l'union de deux singletons distincts).

Proposition 3 (Image directe)

Soit $\mathcal{C}$ un ensemble convexe de $\mathcal{X}$ et $A : \mathcal{X} \to \mathcal{Y}$ une application linéaire. Alors

$$A(\mathcal{C}) = \left\{ A\,x \in \mathcal{Y} \mid x \in \mathcal{C} \right\}$$

est un ensemble convexe de $\mathcal{Y}$.

Démonstration : Soit y_1 et y_2 deux points de $A(\mathcal{C})$. Par définition de l'image directe, il existe donc $(x_1, x_2) \in \mathcal{C}^2$ tel que

$$y_1 = A\,x_1 \qquad \text{et} \qquad y_2 = A\,x_2$$

Soit $\lambda \in [\,0\,;1\,]$. Puisque $\mathcal{C}$ est convexe, on a $\lambda\,x_1 + (1-\lambda)\,x_2 \in \mathcal{C}$. L'application A est linéaire, donc $A\big(\lambda\,x_1 + (1-\lambda)\,x_2\big) = \lambda\,A\,x_1 + (1-\lambda)\,A\,x_2$, ce qui implique que $\lambda\,y_1 + (1-\lambda)\,y_2 \in A(\mathcal{C})$. ■

Proposition 4 (Image réciproque)

Soit $\mathcal{C}$ un ensemble convexe de $\mathcal{Y}$ et $A : \mathcal{X} \to \mathcal{Y}$ une application linéaire. Alors

$$A^{-1}(\mathcal{C}) = \left\{ x \in \mathcal{X} \mid A\,x \in \mathcal{C} \right\}$$

est un ensemble convexe de $\mathcal{X}$.

Démonstration : Soit x_1 et x_2 deux points de $A^{-1}(\mathcal{C})$. Par définition de l'image réciproque, on a

$$(A\,x_1, A\,x_2) \in \mathcal{C}^2$$

Soit $\lambda \in [\,0\,;1\,]$. Par convexité de $\mathcal{C}$, le point $\lambda\,A\,x_1 + (1-\lambda)\,A\,x_2$ appartient à $\mathcal{C}$. L'application A est linéaire donc $\lambda\,A\,x_1 + (1-\lambda)\,A\,x_2 = A\big(\lambda\,x_1 + (1-\lambda)\,x_2\big)$ Il s'ensuit que $\lambda\,x_1 + (1-\lambda)\,x_2 \in A^{-1}(\mathcal{C})$. ■

Exemple

Soit $G : \mathcal{X} \to \mathbb{R}$ une application affine ; il existe donc $(a, b) \in \mathcal{X} \times \mathbb{R}$ tel que

$$\forall\, x \in \mathcal{X}, \qquad G(x) = \langle a, x \rangle + b$$

Alors les ensembles suivants

$$\left\{ x \in \mathcal{X} \mid G(x) = 0 \right\} \qquad \text{et} \qquad \left\{ x \in \mathcal{X} \mid G(x) \leq 0 \right\}$$

sont convexes, car ils peuvent être vus comme les images réciproques par la forme linéaire $x \mapsto \langle a, x \rangle$ du singleton $\{-b\}$ et de l'intervalle ouvert $]\,-\infty\,;-b\,]$, qui sont tous les deux des ensembles convexes de $\mathbb{R}$.

1.3 Projection orthogonale sur un convexe fermé

Définition 4 (Projection orthogonale sur un convexe)

Soit $\mathcal{C} \subset \mathcal{X}$ un ensemble convexe et $x^0 \in \mathcal{X}$. On appelle *projection orthogonale* de x^0 sur $\mathcal{C}$ tout point noté $\mathrm{proj}_{\mathcal{C}}(x^0)$ de $\mathcal{C}$ défini par

$$\forall\, x \in \mathcal{C}, \qquad \|\mathrm{proj}_{\mathcal{C}}(x^0) - x^0\| \leq \|x - x^0\|$$

Remarque : La définition de la projection orthogonale peut être vue comme un problème d'optimisation. Dans le module A5, on généralisera ce concept à travers la notion d'opérateur proximal. Dans le module B8, on considérera des projections définies à l'aide de métriques autre que la norme euclidienne.

Lorsque $\mathcal{C}$ est convexe, fermé et non vide, la projection orthogonale est bien définie :

Théorème 1 (Projection orthgonale sur un ensemble convexe fermé non vide)

Soit $\mathcal{C} \subset \mathcal{X}$ est un ensemble convexe fermé non vide. Alors la projection orthogonale de tout point x_0 sur $\mathcal{C}$ existe et est unique.

Démonstration : Démontrons séparément l'existence et l'unicité de la projection.

- **Existence.** Soit $(x_k)_{k\in\mathbb{N}^*}$ une suite de $\mathcal{C}$ vérifiant

  $$\lim_{k\to+\infty} \left\| x^0 - x_k \right\| = \inf_{x\in\mathcal{C}} \left\| x^0 - x \right\| \in \mathbb{R}^+$$

  Notons α cette quantité. À l'aide de l'identité du parallélogramme, on obtient

  $$4\left\| x^0 - \frac{x_j + x_k}{2} \right\|^2 + \|x_j - x_k\|^2 = 2\left(\left\| x^0 - x_j \right\|^2 + \left\| x^0 - x_k \right\|^2 \right)$$

  L'ensemble $\mathcal{C}$ est convexe et $(x_j, x_k) \in \mathcal{C}^2$ donc $(x_j + x_k)/2 \in \mathcal{C}$. En particulier, par définition de α, on peut minorer le premier terme de cette égalité. Ainsi,

  $$0 \leq \|x_j - x_k\|^2 \leq 2\left(\left\| x^0 - x_j \right\|^2 + \left\| x^0 - x_k \right\|^2 \right) - 4\,\alpha^2$$

  En faisant tendre j et k vers $+\infty$ dans cette relation, le terme de droite convergeant 0, on en déduit par encadrement que la suite $(x_k)_{k\in\mathbb{N}^*}$ est une suite de Cauchy, donc convergente car $\mathcal{X}$ est complet. Sa limite $\mathrm{proj}_{\mathcal{C}}(x^0)$ appartient à $\mathcal{C}$ car ce dernier est fermé. Par continuité, on obtient alors que

  $$\forall\, x \in \mathcal{C}, \qquad \left\| x^0 - x \right\| \geq \alpha = \lim_{k\to+\infty} \left\| x^0 - x_k \right\| = \left\| x^0 - \mathrm{proj}_{\mathcal{C}}(x^0) \right\|$$

- **Unicité.** On suppose qu'il existe deux points $z \neq z' \in \mathcal{C}$ tels que

  $$\forall\, x \in \mathcal{C}, \qquad \|z - x^0\| = \|z' - x^0\| \leq \|x - x^0\|$$

  L'identité du parallélogramme donne :

  $$4\left\| x^0 - \frac{z + z'}{2} \right\|^2 + \left\| z - z' \right\|^2 = 2\left(\left\| x^0 - z \right\|^2 + \left\| x^0 - z' \right\|^2 \right) = 4\left\| x^0 - z \right\|^2$$

  L'ensemble $\mathcal{C}$ est convexe et $(z, z') \in \mathcal{C}^2$ donc $(z + z')/2 \in \mathcal{C}$ et

  $$\frac{1}{2}\left\| x^0 - \frac{z + z'}{2} \right\|^2 < \frac{1}{2}\left\| x^0 - z \right\|^2$$

  car $z \neq z'$, ce qui est absurde. ■

2 Fonctions convexes

2.1 Définition et exemples

Définition 5 (Fonction convexe)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. On dit que f est *convexe* si son domaine $\operatorname{dom} f = \{x \in \mathcal{X} \mid f(x) < +\infty\}$ est convexe et que

$$\forall (x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,], \quad f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Définition 6 (Fonction convexe propre)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**. On dit que f est *propre* si

$$\exists\, x^0 \in \mathcal{X}, \qquad f(x^0) < +\infty$$

Autrement dit, le domaine d'une fonction convexe propre est non vide. On signalera que, chez certains auteurs, les fonctions convexes sont à valeurs dans la droite réelle achevée $\mathbb{R} \cup \{+\infty, -\infty\}$ (voir module B1). Dans ce cas, une fonction convexe f est propre si

$$\exists\, x^0 \in \mathcal{X}, \qquad f(x^0) < +\infty \qquad \text{et} \qquad \forall\, x \in \mathcal{X}, \qquad f(x) > -\infty$$

Notons que, par convexité du domaine de f, celle-ci est équivalente à

$$\forall (x_1, x_2) \in (\operatorname{dom} f)^2, \forall \lambda \in [\,0\,;1\,], \quad f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Ainsi, l'image d'une combinaison convexe par une fonction convexe est inférieure ou égale à la combinaison convexe des images. Géométriquement, dans le cas d'une fonction réelle, cela s'interprète de la manière suivante : la portion du graphe de f dont les abscisses sont compris entre $x_1 \in \mathcal{C}$ et $x_2 \in \mathcal{C}$:

$$\Big\{\big(x, f(x)\big) \in \mathcal{C} \times \mathbb{R} \mid x_1 \leq x \leq x_2\Big\}$$

est située *en-dessous* du segment reliant $\big(x_1, f(x_1)\big)$ et $(x_2, f(x_2))$: on dit que

L'arc est en-dessous de la corde.

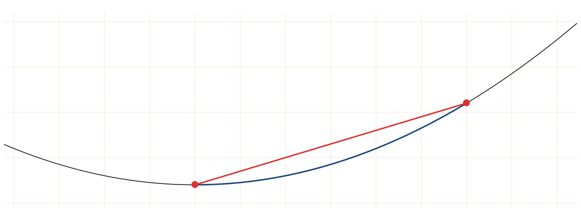

Figure 1 – Exemple du graphe d'une fonction convexe. En rouge l'arc, en bleu l'arc.

Il est aisé de démontrer que toute fonction convexe vérifie l'*inégalité de* Jensen :

Proposition 5 (Inégalité de Jensen)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**. Alors pour tout $(x_i)_{1 \leq i \leq n} \in \mathcal{X}^n$ et pour tout $(\lambda_i)_{1 \leq i \leq n} \in [\,0\,;1\,]^n$, on a

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad \Longrightarrow \qquad f\left(\sum_{i=1}^{n} \lambda_i\, x_i\right) \leq \sum_{i=1}^{n} \lambda_i\, f(x_i)$$

Démonstration : La preuve – par récurrence – est laissée au lecteur.

Définition 7 (Fonction strictement convexe)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. On dit que f est *strictement convexe* si son domaine $\operatorname{dom} f$ est convexe et que

$$\forall\, x_1 \neq x_2 \in \operatorname{dom} f, \forall \lambda \in \,]\,0\,;1\,[\,, \quad f(\lambda\, x_1 + (1-\lambda)\, x_2) < \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Il est évident qu'une fonction strictement convexe est une fonction convexe. La réciproque n'est, quant à elle, généralement pas vraie.

Exemple

Quelques fonctions convexes réelles. Les fonctions suivantes sont convexes :

- les fonctions linéaires et affines ;
- la fonction carrée ;
- la fonction exponentielle.

Exemple

Fonctions affines. Soit $a \in \mathcal{X}$ et $b \in \mathbb{R}$. Considérons la fonction

$$f : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \langle a, x \rangle + b \end{array} \right.$$

Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. Puisque

$$\langle a, \lambda\, x_1 + (1-\lambda)\, x_2 \rangle + b = \lambda\, \big(\langle a, x_1 \rangle + b\big) + (1-\lambda)\, \big(\langle a, x_2 \rangle + b\big)$$

on en déduit que f est convexe.

Proposition 6 (Indicatrice d'un ensemble convexe)

Soit $\mathcal{C} \subset \mathcal{X}$ un ensemble convexe. Alors l'indicatrice de $\mathcal{C}$ définie par

$$\forall\, x \in \mathcal{X}, \qquad \chi_{\mathcal{C}}(x) = \begin{cases} 0 & \text{si } x \in \mathcal{C} \\ +\infty & \text{si } x \notin \mathcal{C} \end{cases}$$

est une fonction convexe.

Démonstration : Il suffit de prouver que $\operatorname{dom} \chi_{\mathcal{C}} = \mathcal{C}$. ■

2.2 Propriétés

Commençons par expliciter le lien entre fonction convexe et ensemble convexe.

Définition 8 (Graphe et épigraphe d'une fonction)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction.

- On définit le *graphe* de f, noté $\operatorname{gr}(f)$:

$$\operatorname{gr}(f) = \Big\{ \big(x, f(x)\big) \in \mathcal{X} \times \mathbb{R} \mid x \in \operatorname{dom} f \Big\}$$

- On définit l'*épigraphe* de f, noté $\operatorname{epi}(f)$:

$$\operatorname{epi}(f) = \Big\{ (x, y) \in \mathcal{X} \times \mathbb{R} \mid x \in \operatorname{dom} f \text{ et } y \geq f(x) \Big\}$$

L'épigraphe est donc la région (fermée) située au-dessus du graphe.

Proposition 7

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Alors on a l'équivalence entre les deux affirmations suivantes :

(i) la fonction f est convexe ;

(ii) son épigraphe $\operatorname{epi}(f)$ est un ensemble convexe de $\mathcal{X} \times \mathbb{R}$.

Démonstration : Montrons séparément les deux sens de l'équivalence.

- **Sens direct.** Soit f une fonction convexe. Soit (x_1, y_1) et (x_2, y_2) deux points de $\operatorname{epi}(f)$. Alors par définition de l'épigraphe, $(x_1, x_2) \in (\operatorname{dom} f)^2$ et $y_1 \geq f(x_1)$ et $y_2 \geq f(x_2)$. Soit $\lambda \in [\,0\,;1\,]$. La fonction f étant convexe, on a

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) = \lambda\, y_1 + (1-\lambda)\, y_2$$

  Ainsi, si $\lambda\, x_1 + (1-\lambda)\, x_2 \in \operatorname{dom} f$, alors $(\lambda\, x_1 + (1-\lambda)\, x_2, \lambda\, y_1 + (1-\lambda)\, y_2)$ est un point de $\operatorname{epi}(f)$.
- **Réciproque.** On suppose que $\operatorname{epi} f$ est convexe. Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. Par définition de l'épigraphe, on a

$$\big(x_1, f(x_1)\big) \in \operatorname{epi} f \qquad \text{et} \qquad \big(x_2, f(x_2)\big) \in \operatorname{epi} f$$

  La convexité de l'épigraphe assure donc que

$$\lambda\,\big(x_1, f(x_1)\big) + (1-\lambda)\,\big(x_2, f(x_2)\big) \in \operatorname{epi} f$$

  c'est-à-dire que $f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$. ■

De la même manière que pour les ensembles convexes, certaines opérations préservent la convexité des fonctions.

Proposition 8 (Combinaison linéaire)

Soit $f_1 : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $f_2 : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions convexes. Alors pour tout $\alpha_1 \geq 0$ et $\alpha_2 \geq 0$, la fonction $\alpha_1\, f_1 + \alpha_2\, f_2$ est convexe.

Démonstration : Laissé au lecteur.

On peut généraliser ce résultat en considérant $A : \mathbb{R}^n \to \mathbb{R}$ une application affine définie par

$$A : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x = (x_{(i)})_{1 \leq i \leq n} & \mapsto & \displaystyle\sum_{i=1}^n a_i\, x_{(i)} + b \end{array} \right.$$

avec $a_i \geq 0$ pour tout $i \in [\![\, 1\,;n\,]\!]$. Si $f = (f_i)_{1 \leq i \leq n}$ est telle que les f_i sont convexes, alors $A \circ f$ est convexe. Pour la composition interne par une application affine, aucune restriction n'est nécessaire quant à la partie linéaire de A, comme le montre la proposition suivante :

Proposition 9 (Composition)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et $A : \mathcal{Y} \to \mathcal{X}$ une application affine. Alors la fonction $f \circ A$ est convexe.

Démonstration : Il suffit d'utiliser la linéarité partielle de A et de remarquer que $b = \lambda\, b + (1-\lambda)\, b$ pour tout $b \in \mathbb{R}$ et tout $\lambda \in [\,0\,;1\,]$. ■

La proposition 9 laisse à penser que toutes les compositions de fonctions convexes ne sont pas nécessairement convexes. Il suffit de considérer par exemple la fonction $-\exp$. De manière générale, si les fonctions en jeu ne sont pas affines, seule la composition externe par une fonction convexe croissante d'une fonction convexe est assurée d'être convexe, comme on peut le voir dans l'exemple suivant.

Exemple

Composition par une fonction convexe croissante. Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe croissante. Alors $g \circ f$ est convexe. En effet, si on suppose que son domaine est non vide (le résultat étant immédiat dans le cas contraire), alors on a pour tout $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$

$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

par convexité de f ; par croissance de g, on a alors

$$g \circ f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq g\big(\lambda\, f(x_1) + (1-\lambda)\, f(x_2)\big)$$

On conclut à l'aide de la convexité de g.

On signale enfin une propriété qui nous sera très utile dans la suite de ce cours :

Proposition 10 (Enveloppe supérieure)

Soit $\mathcal{I} \subset \mathcal{Y}$ et $f_i : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe pour tout $i \in \mathcal{I}$. Alors l'*enveloppe supérieure* de la famille de fonctions $\{f_i\}_{i\in\mathcal{I}}$, définie par

$$\left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & \sup\limits_{i\in\mathcal{I}} f_i(x) \end{array} \right.$$

est une fonction convexe.

Démonstration : Notons f l'enveloppe supérieure des f_i. Soit $x \in \operatorname{dom} f$. Puisque

$$y \geq \sup_{i\in\mathcal{I}} f_i(x) < +\infty \qquad \Longleftrightarrow \qquad \forall\, i \in \mathcal{I}, \quad y \geq f_i(x)$$

on a $(x, y) \in \operatorname{epi} f_i$ pour tout $i \in \mathcal{I}$. Par conséquent,

$$(x, y) \in \operatorname{epi} f \qquad \Longleftrightarrow \qquad (x, y) \in \bigcap_{i\in\mathcal{I}} \operatorname{epi} f_i$$

Les épigraphes des fonctions f_i étant convexes en vertu de la proposition 7, on en déduit que leur intersection l'est également (proposition 2). Aussi, l'épigraphe de f est convexe, ce qui entraîne (proposition 7) que f est convexe. ■

On termine ce paragraphe avec quelques considérations sur la continuité des fonctions convexes en dimension finie. On va admettre le résultat suivant :

Proposition 11

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe telle que $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Soit $x^0 \in \operatorname{int}(\operatorname{dom} f)$. Alors f est continue en x^0.

Démonstration : Admis.

Il est important de considérer des points situés à l'intérieur du domaine, comme le démontre le contre-exemple suivant :

Contre-exemple

Discontinuité au bord. Considérons la fonction $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ définie par

$$\forall\, x \in \mathbb{R}, \qquad f(x) = \begin{cases} 0 & \text{si } x > 0 \\ 1 & \text{si } x = 0 \\ +\infty & \text{si } x < 0 \end{cases}$$

Le graphe de cette fonction – discontinue en 0 – est montré en figure 2. Vérifions qu'elle est convexe. L'inégalité définissant la convexité étant immédiatement vérifiée pour deux points x_1 et x_2 strictement positifs, il reste le cas où $x_1 = 0$ et $x_2 > 0$. Puisque pour tout $\lambda \in \,]\,0\,;1\,[$, on a $\lambda\, x_1 + (1-\lambda)\, x_2 > 0$, et que

$$f\big((\lambda\, x_1 + (1-\lambda)\, x_2\big) = 0 \leq \lambda = \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

on en déduit le résultat désiré.

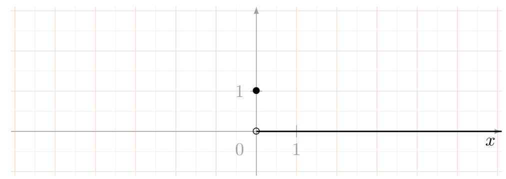

Figure 2 – Cas d'une fonction convexe discontinue au bord de son domaine.

Une conséquence directe de la proposition 10 est qu'une fonction convexe de domaine ouvert est continue sur son domaine. Lorsque son domaine n'est pas ouvert, il est possible de lui associer une fonction convexe de même domaine, en définissant :

$$\forall\, x \in \mathcal{X}, \qquad \underline{f}(x) = \liminf_{y\to x} f(y)$$

Cette fonction est appelée *fermeture* de f. Cette fonction possède la propriété d'être *semi-continue inférieurement*, une propriété plus faible que la continuité mais peut se révéler utile dans certains contextes comme celle de la minimisation.

2.3 Cas différentiable

Si la fonction est différentiable, la convexité est caractérisée grâce au gradient :

Proposition 12

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction *différentiable*. Alors on a équivalence entre les affirmations suivantes :

(i) f est convexe ;

(ii) $\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle$

(iii) $\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1 \rangle \geq 0$

Démonstration : On démontre successivement les différentes implications.

- (i) ⇒ (ii). Supposons que f est convexe. Soit $\lambda \in [\,0\,;1\,]$ et $(x_1, x_2) \in \mathcal{X}^2$ tels que $x_1 \neq x_2$. La définition de la convexité s'écrit

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

soit, en développant les produits et en réarrangeant les termes obtenus,

$$f\big(x_2 + \lambda\,(x_1 - x_2)\big) - f(x_2) \leq \lambda\,\big(f(x_1) - f(x_2)\big)$$

Divisons par $\lambda > 0$ puis faisons tendre λ vers 0. Puisque f est différentiable, la limite suivante existe et vaut

$$\lim_{\lambda\to 0} \frac{f\big(x_2 + \lambda\,(x_1 - x_2)\big) - f(x_2)}{\lambda} = \langle \nabla f(x_2), x_1 - x_2 \rangle$$

- (ii) $\Rightarrow$ (iii). On suppose que (ii) est vérifiée. Soit $(x_1, x_2) \in \mathcal{X}^2$. En particulier, l'inégalité (ii) est vraie pour (x_1, x_2) et (x_2, x_1). En sommant les deux inégalités obtenues, il en découle que

$$f(x_1) + f(x_2) \geq f(x_1) + f(x_2) + \langle \nabla f(x_1) - \nabla f(x_2), x_2 - x_1 \rangle$$

En simplifiant cette relation, on obtient le résultat désiré.
- (iii) $\Rightarrow$ (ii). On suppose que (iii) est vérifiée. On définit

$$g : \begin{cases} [\,0\,;1\,] & \to \mathbb{R} \\ t & \mapsto f\big(x_1 + t(x_2 - x_1)\big) \end{cases}$$

La fonction g est continue sur $[\,0\,;1\,]$ (par composition) et dérivable sur $]\,0\,;1\,[$ (par composition) de dérivée

$$\forall\, t \in \,]\,0\,;1\,[\qquad g'(t) = \langle \nabla f(x_1 + t(x_2 - x_1)), x_2 - x_1 \rangle$$

D'après le théorème des accroissements finis il existe $t \in \,]\,0\,;1\,[$ tel que

$$f\big(x_1 + (x_2 - x_1)\big) - f(x_1) = \big\langle \nabla f\big(x_1 + t\,(x_2 - x_1)\big), x_2 - x_1 \big\rangle$$

En simplifiant le membre de gauche, il devient $f(x_2) - f(x_1)$, tandis que le membre de droite peut être minoré en appliquant l'hypothèse (iii) aux points $x_1 + t\,(x_2 - x_1)$ et x_1 :

$$\Big\langle \nabla f\big(x_1 + t\,(x_2 - x_1)\big) - \nabla f(x_1), \big(x_1 + t\,(x_2 - x_1)\big) - x_1 \Big\rangle \geq 0$$

En simplifiant l'expression précédente, en divisant par t et en faisant tendre t vers 0, on obtient le résultat désiré.
- (ii) $\Rightarrow$ (i). On suppose que (ii) est vérifiée. Soit $(x_1, x_2) \in \mathcal{X}^2$ tel que $x_1 \neq x_2$. Soit $\lambda \in \,]\,0\,;1\,[$. Posons $x_\lambda = (1-\lambda)x_1 + \lambda\, x_2$. On a alors $x_\lambda \neq x_1$ et $x_\lambda \neq x_2$. Appliquons l'hypothèse (ii) aux points (x_1, x_λ) et (x_2, x_λ) :

$$f(x_1) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), x_1 - x_\lambda \rangle \qquad \text{et} \qquad f(x_2) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), x_2 - x_\lambda \rangle$$

On remarque que $x_1 - x_\lambda = \lambda(x_1 - x_2)$ et $x_2 - x_\lambda = (1-\lambda)(x_2 - x_1)$. Multiplions alors la première inégalité par $(1-\lambda)$ et la seconde par λ, les deux facteurs étant positifs. En sommant ces deux inégalités, on obtient alors, après réarrangement, que

$$f(x_\lambda) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

d'où l'on déduit que f est convexe. ■

La proposition 12 assure que, si f est convexe et différentiable, alors, pour $x^0 \in \mathcal{X}$, on a

$$\forall\, x \in \mathcal{X}, \qquad f(x) \geq \langle \nabla f(x^0), x \rangle + f(x^0) - \langle \nabla f(x^0), x^0 \rangle$$

Autrement dit, f est minorée par une fonction affine. Ce genre de fonctions est appelé *minorante affine*.

Définition 9 (Minorante affine)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. On appelle *minorante affine* de f toute fonction affine $g : \mathcal{X} \to \mathbb{R}$ telle que

$$\forall\, x \in \mathcal{X}, \qquad g(x) \leq f(x)$$

Géométriquement, une minorante affine est donc un hyperplan situé en-dessous du graphe de la fonction. On peut généraliser la remarque faite ci-dessus pour toute fonction convexe et propre :

Proposition 13

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et propre. Alors f admet une minorante affine.

Démonstration : Admis.

> On verra dans le module A2 qu'il est possible de généraliser la caractérisation de la proposition 12 dans le cas non différentiable.

3 Forte convexité

3.1 Définition et exemples

Définition 10 (Fonction fortement convexe)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction et soit $\alpha > 0$. On dit que f est *fortement convexe* de module α si son domaine $\operatorname{dom} f$ est convexe et que

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,],$$
$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\, \lambda\, (1-\lambda)\, \|x_1 - x_2\|^2$$

On dit parfois que f est α-convexe. Il est aisé de vérifier qu'une fonction fortement convexe est strictement convexe :

Proposition 14

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction fortement convexe de module α. Alors f est une fonction strictement convexe.

Démonstration : Laissé au lecteur.

Exemple

Fonction fortement convexe. Soit $x^0 \in \mathcal{X}$ et $\alpha > 0$. Considérons la fonction suivante :

$$f : \begin{cases} \mathcal{X} & \to \quad \mathbb{R} \\ x & \mapsto \quad \dfrac{\alpha}{2}\, \|x - x^0\|^2 \end{cases}$$

Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. En remarquant que $x^0 = \lambda\, x^0 + (1-\lambda)\, x^0$ et en développant le premier carré, on obtient que

$$\frac{\alpha}{2}\left\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\right\|^2 - \frac{\alpha}{2}\,\lambda\,\|x_1 - x^0\|^2 - \frac{\alpha}{2}\,(1-\lambda)\,\|x_2 - x^0\|^2$$
$$= -\frac{\alpha}{2}\,\lambda\,(1-\lambda)\big(\|x_1 - x^0\|^2 + \|x_2 - x^0\|^2 - 2\,\langle x_1 - x^0, x_2 - x^0\rangle\big)$$

En utilisant une identité remarquable pour simplifier le terme entre parenthèses dans l'expression précédente, on prouve que f est une fonction fortement convexe de module α.

3.2 Propriétés

Proposition 15

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Alors on a l'équivalence entre les affirmations suivantes :

(i) f est fortement convexe de module α ;

(ii) pour tout $x^0 \in \mathcal{X}$, la fonction

$$x \mapsto f(x) - \frac{\alpha}{2}\,\|x - x^0\|^2$$

est une fonction convexe.

Démonstration : Soit $x^0 \in \mathcal{X}$. Le calcul effectué dans l'exemple précédent prouve que, pour tout $(x_1, x_2) \in \mathcal{X}^2$ et tout $\lambda \in [\,0\,;1\,]$, on a

$$\left\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\right\|^2 - \lambda\,\|x_1 - x^0\|^2 - (1-\lambda)\,\|x_2 - x^0\|^2 = -\lambda\,(1-\lambda)\,\|x_1 - x_2\|^2$$

Par définition, f est fortement convexe de module α si et seulement si

$$\forall\,(x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,],$$
$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \le \lambda\, f(x_1) + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\,\lambda\,(1-\lambda)\,\|x_1 - x_2\|^2$$

soit, en multipliant par $\alpha/2$,

$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) - \frac{\alpha}{2}\left\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\right\|^2$$
$$\le \lambda\, f(x_1) - \frac{\alpha}{2}\,\lambda\,\|x_1 - x^0\|^2 + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\,(1-\lambda)\,\|x_2 - x^0\|^2$$

qui n'est autre que la définition de la convexe appliquée à la fonction apparaissant dans l'affirmation (ii). ■

En posant g la fonction apparaissant dans l'affirmation (ii) de la proposition 15, on assure l'existence d'une fonction convexe $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ telle que

$$\forall\, x \in \mathcal{X}, \qquad f(x) = g(x) + \frac{\alpha}{2}\,\|x - x^0\|^2$$

Cette même proposition permet de caractériser les fonctions fortement convexes différentiables à l'aide de leur gradient :

Corollaire 1

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction différentiable. Alors on a équivalence entre les affirmations suivantes :

(i) f est fortement convexe de module α ;

(ii) $\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \ge f(x_1) + \langle \nabla f(x_1), x_2 - x_1\rangle + \frac{\alpha}{2}\,\|x_1 - x_2\|^2$

(iii) $\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \ge \alpha\,\|x_1 - x_2\|^2$

Démonstration : Il suffit d'appliquer la proposition 12 à la fonction

$$x \mapsto f(x) - \frac{\alpha}{2}\,\|x - x_1\|^2$$

qui est une fonction différentiable par hypothèse. ■

> On verra dans le module A2 qu'il est possible de généraliser cette caractérisation dans le cas continu non différentiable.

Terminons ce module avec un résultat justifiant l'intérêt que l'on peut porter à la forte convexité en optimisation :

Proposition 16

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. On suppose que J est fortement convexe, de module α, et propre. On suppose également que $\operatorname{dom} J$ est fermé et que J est continue sur $\operatorname{dom} J$. Alors il existe un unique point $x^* \in \mathcal{X}$ tel que

$$\forall\, x \in \mathcal{X}, \qquad J(x) \ge J(x^*)$$

et, par ailleurs, on a

$$\forall\, x \in \mathcal{X}, \qquad \|x - x^*\|^2 \le \frac{4}{\alpha}\big(J(x) - J(x^*)\big)$$

Pour la définition d'une fonction s.c.i., on se reportera au module A3.

Démonstration : Considérons séparément les deux affirmations de la proposition.

- **Existence et unicité du minimiseur x^*.** Admis pour l'instant (démontré au module B1). Il s'agit d'une propriété commune à toute fonction fortement convexe.
- **Majoration de la distance au minimiseur.** On écrit la définition de la forte convexité aux points x et x^* :

$$\forall\,\lambda \in [\,0\,;1\,], \quad J(\lambda\, x + (1-\lambda)\, x^*) \le \lambda\, J(x) + (1-\lambda)\, J(x^*) - \frac{\alpha}{2}\,\lambda\,(1-\lambda)\,\|x - x^*\|^2$$

Par définition de x^*, on a

$$\forall\,\lambda \in [\,0\,;1\,], \quad J(x^*) \le J(\lambda\, x + (1-\lambda)\, x^*)$$

de sorte que l'inégalité précédente se simplifie en

$$\forall \lambda \in [\,0\,;1\,]\,, \quad \|x - x^*\|^2 \leq \frac{2}{\alpha\,(1-\lambda)}\big(J(x) - J(x^*)\big)$$

Il suffit alors de majorer la quantité $1 - \lambda$ pour obtenir le résultat désiré. ■

Une conséquence importante de ce dernier résultat est la suivante. On suppose que J vérifie les hypothèses de la proposition précédente. Si $(x_k)_{k\in\mathbb{N}}$ est une suite de $\mathcal{X}$ telle que

$$\lim_{k\to+\infty} J(x_k) = \inf_{x\in\mathcal{X}} J(x)$$

(on dit que $(x_k)_{k\in\mathbb{N}}$ est une suite minimisante de J, et on notera qu'elle converge vers le minimum de J, qui vaut $J(x^*)$), alors

$$\lim_{k\to+\infty} x_k = x^*$$

Autrement dit, toute suite minimisante de J converge vers le minimiseur de J. Ce résultat n'est évident pas vrai en général (il suffit de considérer par exemple une fonction constante sur $\mathbb{R}$ et la suite des entiers naturels).

Pour aller plus loin

Enveloppe supérieure d'une famille de fonctions convexes. Cette notion sera utilisée dans le contexte de la dualité min-max (module A6). En particulier, on verra qu'il est possible de représenter toute fonction convexe comme l'enveloppe supérieure d'une famille de fonctions affines, ce qui conduira à la notion de conjuguée convexe (module A8).

Généralisation de la notion de différentiabilité. Le gradient d'une fonction convexe possède des propriétés remarquables, que l'on préservera en définissant une généralisation de la différentielle aux fonctions convexes non différentiables, ou encore non convexes et non différentiables (module A2). Cette généralisation, appelée *sous-différentiel*, se révélera en particulier très utile pour caractériser les minima globaux d'une fonction convexe (module B1). On verra par ailleurs qu'un des intérêts de l'optimisation convexe réside dans l'absence de minima locaux.

Forte convexité. La forte convexité est une propriété intéressante pour l'optimisation, car elle garantit à la fois l'existence et l'unicité du minimiseur (module B1). De plus, elle permet d'assurer des comportements plus intéressants pour les algorithmes d'optimisation, tant en termes de stabilité qu'en termes de vitesse de convergence (module B1). On verra en outre que la forte convexité est une forme de régularité dans le domaine dual (module A8).