Module B1

Algorithmes d'optimisation du premier ordre

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. Pour toutes les autres notations et définitions, le lecteur est invité à se reporter aux modules précédents.

1 Problèmes d'optimisation

1.1 Droite réelle achevée

Définition 1 (Droite réelle achevée)

On appelle *droite réelle achevée* l'ensemble suivant, noté $\overline{\mathbb{R}}$:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\} = [\,-\infty\,;+\infty\,]$$

En adjoignant à la droite réelle les deux valeurs infinies $+\infty$ et $-\infty$, on est amené à définir de nouvelles règles de calculs. Certaines sont bien définies : on commence par exemple avec la somme et le produit avec un scalaire (non nul pour le produit) :

$$\forall\,(\alpha,\beta) \in \mathbb{R}^* \times \mathbb{R}, \qquad \alpha \times (+\infty) + \beta = \alpha \times (+\infty)$$

où, pour tout $\alpha \neq 0$,
$$\alpha \times (+\infty) = \begin{cases} +\infty & \text{si } \alpha > 0 \\ -\infty & \text{si } \alpha < 0 \end{cases}$$

Le cas du produit avec 0 est plus délicat ; on songera par exemple aux fonctions $f : x \mapsto x$ et $g : x \mapsto 1/x$, pour lesquelles on a

$$f(0^+) = 0, \qquad g(0^+) = +\infty \qquad \text{et} \qquad f(0^+) \times g(0^+) = 1$$

car $f \times g = 1$, alors que

$$f(0^+) = 0, \qquad \big(g(0^+)\big)^2 = +\infty \qquad \text{et} \qquad f(0^+) \times g(0^+) = +\infty$$

car $(f \times g^2)(x) = 1/x$ quel que soit $x \in \mathbb{R}$. On voit donc qu'il n'est pas raisonnable d'attribuer une valeur unique à ce produit. Concernant les opérations entre valeurs infinies, on a d'une part

$$(+\infty) + (+\infty) = +\infty \qquad \text{et} \qquad (-\infty) + (-\infty) = -\infty$$

et d'autre part

$$(+\infty) \times (+\infty) = (-\infty) \times (-\infty) = +\infty \qquad \text{et} \qquad (+\infty) \times (-\infty) = (-\infty) \times (+\infty) = -\infty$$

Il n'est pas non plus raisonnable de définir la somme $(+\infty) + (-\infty)$ (ou, de manière équivalente, $+\infty - (+\infty)$) et son opposé ; pour s'en convaincre, il suffit de considérer la fonction $g : x \mapsto 1/x$, pour laquelle on a par exemple

$$g(0^+) - g(0^+) = 0 \qquad \text{et} \qquad \big(g(0^+)\big)^2 - g(0^+) = +\infty$$

(car $g(x) - g(x) = 0$ et $(g(x))^2 - g(x) = x^2 - x$). On peut résumer ces observations en *définissant* les conventions suivantes :

Définition 2

On peut étendre la définition de la somme et du produit usuels dans $\mathbb{R}$ sur la droite réelle achevée en posant pour la somme

$$\forall\,\beta \in \,]-\infty\,;+\infty\,], \qquad (+\infty) + \beta = +\infty$$
$$\forall\,\beta \in [\,-\infty\,;+\infty\,[, \qquad (-\infty) - \beta = -\infty$$

et pour le produit

$$\forall\,\alpha \in \,]\,0\,;+\infty\,], \qquad \alpha \times (+\infty) = +\infty \qquad \text{et} \qquad \alpha \times (-\infty) = -\infty$$
$$\forall\,\alpha \in [\,-\infty\,;0\,[, \qquad \alpha \times (+\infty) = -\infty \qquad \text{et} \qquad \alpha \times (-\infty) = +\infty$$

ces deux lois de composition internes restant commutatives.

> Il n'est donc pas possible de munir de manière raisonnable la droite réelle achevée d'une structure de groupe ou de corps . En revanche, l'ensemble $\mathbb{R} \cup \{+\infty\}$, muni de la somme est un *monoïde commutative*, c'est-à-dire que la somme est une loi de composition interne associative et commutative, d'élément neutre 0.

D'autres opérations sur la droite réelle achevée nous seront utiles dans ce cours. Il s'agit tout d'abord de la relation d'ordre

$$\forall\,\alpha \in \mathbb{R}, \qquad -\infty < \alpha < +\infty$$

de sorte que $+\infty$ (resp. $-\infty$) est le plus grand (resp. le plus petit) élément de $\overline{\mathbb{R}}$. Tout ensemble de la droite réelle achevée admet une borne supérieure et une borne inférieure dans $\mathbb{R} \cup \{+\infty, -\infty\}$; plus précisément, si $\mathcal{E} \subset \mathbb{R} \cup \{+\infty\}$ (resp. $\mathcal{E} \subset \mathbb{R} \cup \{-\infty\}$) est non vide, alors

$$\sup \mathcal{E} \in \mathbb{R} \cup \{+\infty\} \qquad (\text{resp.} \qquad \inf \mathcal{E} \in \mathbb{R} \cup \{-\infty\})$$

tandis que $\sup \emptyset = \sup\{-\infty\} = -\infty$ et $\inf \emptyset = \inf\{+\infty\} = +\infty$

On rappelle que la borne inférieure d'un ensemble $\mathcal{E} \subset \overline{\mathbb{R}}$ est définie comme le plus grand minorant de $\mathcal{E}$. Lorsqu'elle est finie, elle est caractérisée de la manière suivante :

$$\forall\,\varepsilon > 0, \exists\, x \in \mathcal{E}, \qquad m \leq x \leq m + \varepsilon$$

et lorsqu'elle vaut $-\infty$, elle est caractérisée par

$$\forall\, m \in \mathbb{R}, \exists\, x \in \mathcal{E}, \qquad x \leq m$$

La borne supérieure d'un ensemble de $\overline{\mathbb{R}}$ est définie de manière analogue, en inversant le sens de l'inégalité. On définit alors les bornes supérieure et inférieure de f sur un ensemble $\mathcal{E} \subset \mathcal{X}$ comme

$$\sup_{x\in\mathcal{E}} f(x) = \sup \left\{ f(x) \mid x \in \mathcal{E} \right\} \qquad \text{et} \qquad \inf_{x\in\mathcal{E}} f(x) = \inf \left\{ f(x) \mid x \in \mathcal{E} \right\}$$

et on dit que le minimum (resp. le maximum) de f sur $\mathcal{E}$ est atteint lorsque la borne inférieure (resp. supérieure) de f sur $\mathcal{E}$ est *finie* et qu'il existe un élément de $\mathcal{E}$ pour lequel f prend cette valeur.

Attention, dans le cas des fonctions à valeurs sur la droite réelle achevée, la caractérisation séquentielle de la continuité en un point s'écrit de la manière suivante :

Définition 3 (Fonction continue)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $x^0 \in \operatorname{dom} f$. On dit que f est continue en x^0 si pour toute suite $(x_k)_{k\in\mathbb{N}}$ de $\operatorname{dom} f$, on a

$$\lim_{k\to+\infty} x_k = x^0 \qquad \lim_{k\to+\infty} f(x_k) = f(x^0)$$

Si $\operatorname{dom} f \neq \mathcal{X}$, on dit que f est *continue sur son domaine* si f est continue en tout point de son domaine. On dit que f est *continue* si $\operatorname{dom} f = \mathcal{X}$ et que f est continue sur son domaine.

1.2 Problème d'optimisation

Résoudre un problème d'optimisation, c'est chercher à minimiser ou à maximiser une fonction dite *objectif*.

Définition 4 (Minimiseurs et minimum)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction et $x^* \in \mathcal{X}$ tel que $J(x^*) \in \mathbb{R}$. On dit que x^* est un *minimiseur (global)* de J si

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

L'ensemble des minimiseurs (globaux) de J (éventuellement vide) est noté

$$\operatorname*{argmin}_{x\in\mathcal{X}} J(x) \subset \mathcal{X}$$

Dans ce cas, la valeur réelle $J(x^*)$ est appelée *minimum (global)* de J, notée

$$\min_{x\in\mathcal{X}} J(x) \in \mathbb{R}$$

Notons que, si le minimum (lorsqu'il existe) d'une fonction est unique, ses minimiseurs peuvent ne pas l'être (il suffit de considérer n'importe quelle fonction constante). En particulier, tous les minimiseurs de J prennent la même valeur par J, qui vaut le minimum de J :

$$\forall\, x^* \in \operatorname*{argmin}_{x\in\mathcal{X}} J(x), \qquad J(x^*) = \min_{x\in\mathcal{X}} J(x)$$

Définition 5 (Maximiseurs et maximum)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction et $x^* \in \mathcal{X}$ tel que $J(x^*) \in \mathbb{R}$. On dit que x^* est un *maximiseur (global)* de J si

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

L'ensemble des maximiseurs (globaux) de J (éventuellement vide) est noté

$$\operatorname*{argmax}_{x\in\mathcal{X}} J(x) \subset \mathcal{X}$$

Dans ce cas, la valeur réelle $J(x^*)$ est appelée *maximum (global)* de J, notée

$$\max_{x\in\mathcal{X}} J(x) \in \mathbb{R}$$

Ainsi, x^* est un maximiseur de J si et seulement si x^* est un minimiseur de $-J$. C'est pourquoi on peut confondre la notion d'optimisation et celle de minimisation, comme on le fera à partir de maintenant.

On précise parfois minimiseur / maximiseur / minimum / maximum global pour les distinguer de leur homologue local :

Définition 6 (Minimiseurs locaux)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction et $x^* \in \mathcal{X}$ tel que $J(x^*) \in \mathbb{R}$. On dit que x^* est un *minimiseur local* de J s'il existe un voisinage $\mathcal{V}(x^*) \subset \mathcal{X}$ tel que

$$\forall\, x \in \mathcal{V}(x^*), \qquad J(x^*) \leq J(x)$$

Dans ce cas, la valeur réelle $J(x^*)$ est appelée *minimum local* de J.

On peut définir de manière analogue les maximiseurs et maxima locaux. Contrairement au minimum / maximum global, une fonction peut admettre plusieurs minima / maxima locaux.

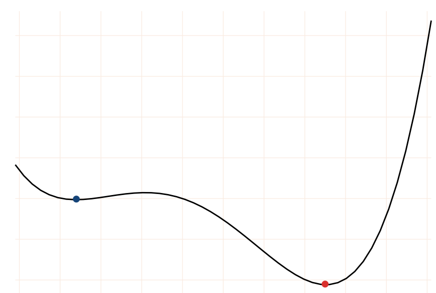

FIGURE 1 – Minimiseur global *versus* minimiseur local. En bleu un minimiseur local, en rouge le minimiseur global.

Il est clair qu'un minimiseur (resp. minimum, maximiseur, maximum) global est un minimiseur (resp. minimum, maximiseur, maximum) local. En général, la réciproque est fausse, sauf dans le cas convexe :

Proposition 1

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**. Si $x^* \in \mathcal{X}$ est un minimiseur local de J, alors x^* est un minimiseur global de J.

Démonstration : Soit $x^* \in \mathcal{X}$ un minimiseur local de J, c'est-à-dire qu'il existe un ouvert $\mathcal{V}(x^*)$ contenant x^* tel que

$$\forall\, x \in \mathcal{V}(x^*), \qquad J(x^*) \leq J(x)$$

Supposons que x n'est pas un minimiseur global de J ; il existe donc $z \in \mathcal{X}$ tel que

$$J(z) < J(x^*)$$

Soit $\lambda \in \,]\,0\,;1\,[$. Posons $z_\lambda = \lambda\, z + (1-\lambda)\, z$. On a alors

$$\|z_\lambda - z\| = \lambda\, \|z - x\|$$

En particulier, si λ est suffisamment petit, alors $z_\lambda \in \mathcal{V}(x^*)$. Par conséquent, si λ est choisi suffisamment petit, alors $J(z_\lambda) \geq J(x^*)$. Par ailleurs, J est convexe donc

$$J(z_\lambda) = J(\lambda\, z + (1-\lambda)\, x^*) \leq \lambda\, J(z) + (1-\lambda)\, J(x^*)$$

On en déduit que, après simplification,

$$\lambda\, J(x^*) \leq \lambda\, J(z)$$

Il suffit alors de diviser par λ, qui est strictement positif, pour établir une contradiction avec l'hypothèse initiale. ■

On peut à présent formaliser davantage le cadre des problèmes d'optimisation. On appelle *problème d'optimisation* la recherche d'un minimiseur ou d'un maximiseur d'une fonction objectif $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$, c'est-à-dire, dans le cas d'un problème de minimisation, la recherche d'un point $x^* \in \mathcal{X}$ tel que $J(x^*) \in \mathbb{R}$ et

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

On le note de manière équivalente

$$\text{Minimiser } J \qquad \text{ou} \qquad \min_{x \in \mathcal{X}} J(x) \tag{$\mathcal{P}$}$$

On parle de problème d'optimisation *sans contraintes*, car on ne contraint pas explicitement x à appartenir à un ensemble particulier. *A contrario*, un *problème d'optimisation sous contraintes* la recherche d'un point $x^* \in \mathcal{C}$ tel que $J(x^*) \in \mathbb{R}$ et que

$$\forall\, x \in \mathcal{C}, \qquad J(x^*) \leq J(x)$$

où l'ensemble $\mathcal{C} \subset \mathcal{X}$ est appelé *ensemble des contraintes* ou *ensemble admissible* (dans ce cas, tout point $x \in \mathcal{C}$ est appelé *point admissible*). Ce problème est alors noté

$$\text{Minimiser } J(x) \text{ sous les contraintes } x \in \mathcal{C} \qquad \text{ou} \qquad \min_{x \in \mathcal{C}} J(x) \tag{$\mathcal{P}_{\rm c}$}$$

En utilisant une fonction indicatrice, il est toujours possible d'écrire un problème d'optimisation sous contraintes comme un problème d'optimisation non contraint, en définissant une nouvelle fonction objectif $\widetilde{J} : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ en posant

$$\forall\, x \in \mathcal{X}, \qquad \widetilde{J}(x) = J(x) + \chi_{\mathcal{C}}(x)$$

C'est la raison pour laquelle, sauf dans le cas particulier des problèmes d'optimisation convexe lisse sous contraintes d'inégalités et d'égalités (module A8), on ne considérera dans ce cours que des problèmes d'optimisation non contrainte.

Terminons ce paragraphe avec la notion de problèmes équivalents. Cette notion est formalisée ici d'une manière non standard :

Définition 7 (Problèmes d'optimisation équivalents)

Soit $J_1 : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ et $J_2 : \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ deux fonctions. On note $(\mathcal{P}_1)$ (resp. $(\mathcal{P}_2)$) le problème d'optimisation de fonction objectif J_1 (resp. J_2). On dit que les problèmes $(\mathcal{P}_1)$ et $(\mathcal{P}_2)$ sont équivalents s'il existe deux applications $M_1 : \mathcal{X} \to 2^{\mathcal{Z}}$ et $M_2 : \mathcal{Z} \to 2^{\mathcal{X}}$ telles que

$$x^* \in \operatorname*{argmin}_{x \in \mathcal{X}} J_1(x) \qquad \Longrightarrow \qquad M_1(x^*) \in \operatorname*{argmin}_{z \in \mathcal{Z}} J_2(z)$$

et

$$z^* \in \operatorname*{argmin}_{z \in \mathcal{Z}} J_2(z) \qquad \Longrightarrow \qquad M_2(z^*) \in \operatorname*{argmin}_{x \in \mathcal{X}} J_1(x)$$

Un cas particulier de problèmes équivalents est celui où les fonctions objectifs partagent mêmes minimiseurs. Ainsi, cette définition permet de généraliser cette notion ; il faut noter notamment que deux problèmes équivalents selon cette définition n'ont pas nécessairement le même nombre de solutions. Même lorsqu'elle n'est pas mentionnée de manière formelle, la notion de problèmes équivalents est très utile pour la résolution de problèmes d'optimisation. Il est parfois possible d'introduire un problème auxiliaire dont les solutions permettent d'obtenir les solutions du problème initialement considéré. Remarquons enfin qu'aucune condition n'est imposée sur le minimum atteint dans les problèmes considérés.

Exemple

Problèmes d'optimisation équivalents. Considérons les deux fonctions objectifs suivantes :

$$J_1 : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & -x^2 + \chi_{[\,0\,;1\,]} \end{array} \right. \qquad \text{et} \qquad J_2 : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \cup \{+\infty\} \\ z & \mapsto & -z^2 + \chi_{[\,-1\,;1\,]} \end{array} \right.$$

Alors l'optimisation de ces deux fonctions définit deux problèmes équivalents, avec

$$M_1(x) = \{x, -x\} \qquad \text{et} \qquad M_2(z) = \{|z|\}$$

On remarque en particulier que la première fonction admet un unique minimiseur, tandis que le second en admet deux.

1.3 Existence et unicité des minimiseurs

Commençons par noter qu'une condition nécessaire d'existence de minimiseurs est le fait que J ne prenne pas la valeur $-\infty$ et ne soit pas identiquement égale à $+\infty$. Si J est convexe, cela revient à imposer que J soit propre.

Une autre notion qui apparaît souvent en optimisation est la suivante :

Définition 8 (Suite minimisante)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$. On appelle *suite minimisante* de J toute suite $(x_k)_{k\in\mathbb{N}}$ de $\mathcal{X}$ vérifiant

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

Notons que, si $\operatorname{dom} J \neq \emptyset$, alors de toute suite minimisante, on peut extraire une sous-suite d'éléments appartenant au domaine de J. Dans le cas contraire, les $J(x_k)$ valent $+\infty$ à partir d'un certain rang, ce qui contredit le fait que la suite $(J(x_k))_{k\in\mathbb{N}}$ tend vers la borne inférieure de J (qui appartient à $\mathbb{R} \cup \{-\infty\}$ si le domaine de J est non vide). Par conséquent, on considère dorénavant que les suites minimisantes vivent dans le domaine de J.

Proposition 2

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$. On suppose que $\operatorname{dom} J \neq \emptyset$. Alors J admet une suite minimisante.

Démonstration : La preuve repose sur la caractérisation de la borne inférieure d'un sous-ensemble de $\overline{\mathbb{R}}$. Notons

$$m = \inf_{x\in\mathcal{X}} J(x) \in \mathbb{R} \cup \{-\infty\}$$

- **Cas** $m = -\infty$**.** Pour tout $k \in \mathbb{N}$, il existe $x_k \in \operatorname{dom} J$ tel que $J(x_k) \leq -k$. Par comparaison,
$$\lim_{k\to+\infty} J(x_k) = -\infty$$
- **Cas** $m \in \mathbb{R}$**.** Pour tout $k \in \mathbb{N}$, si l'on pose $\varepsilon_k = 1/(k+1)$, alors il existe $x_k \in \operatorname{dom} J$ tel que
$$m \leq J(x_k) < m + \frac{1}{k+1}$$
de sorte que, par encadrement, on a
$$\lim_{k\to+\infty} J(x_k) = m \quad \blacksquare$$

Lorsque l'on étudie un problème d'optimisation, une question importante est celle de l'existence d'un minimiseur. Les résultats suivants permettent d'établir l'existence de minimiseurs sous certaines hypothèses. Si la fonction objectif considérée n'entre pas dans le champ d'application de ces résultats, elle peut ne pas admettre de minimiseur. Dans ce cas, pour le démontrer, une manière de procéder est d'exhiber une suite minimisante telle que la suite des images par la fonction objectif tend vers $-\infty$.

On commence par mentionner quelques résultats d'existence de minimiseurs. Il ne s'agit bien évidemment pas d'une liste exhaustive, mais de résultats classiques et généralement suffisants dans les applications. Le résultat suivant est une généralisation du théorème bien connu de Weierstrass :

Proposition 3 (Existence de minimiseurs I)

Soit $f : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ une fonction. On suppose que son domaine est compact non vide et que f est continue sur son domaine. Alors f admet un minimiseur.

Démonstration : Admis.

Attention, la continuité est essentielle :

Contre-exemple

Fonction discontinue sur un compact. Considérons la fonction définie par :

$$\forall\, x \in \mathbb{R}, \qquad J(x) = \begin{cases} 1 & \text{si } x = 0 \\ x^2 & \text{si } x \in [\,-1\,;0\,[\,\cup\,]\,0\,;1\,] \\ +\infty & \text{sinon} \end{cases}$$

Il est clair que cette fonction n'est pas continue en 0, mais que son domaine est compact. Alors on peut vérifier que

$$\forall\, x \in \mathbb{R}, \qquad J(x) > 0 \qquad \text{et} \qquad \inf_{x\in\mathbb{R}} J(x) = 0$$

Autrement dit, J n'admet pas de minimiseur. Pourtant, cette fonction est coercive (cf. définition plus bas).

De la même manière, le caractère compact (c'est-à-dire borné et fermé en dimension finie) du domaine est essentiel :

Contre-exemple

Domaine non fermé. Considérons la fonction définie par :

$$\forall\, x \in \mathbb{R}, \qquad J(x) = \begin{cases} x & \text{si } x \in \,]\,-1\,;1\,[\\ +\infty & \text{sinon} \end{cases}$$

Alors on peut vérifier que

$$\forall\, x \in \mathbb{R}, \qquad J(x) > -1 \qquad \text{et} \qquad \inf_{x\in\mathbb{R}} J(x) = -1$$

Autrement dit, J n'admet pas de minimiseur. Pourtant, cette fonction est coercive (cf. définition plus bas), convexe et continue sur son domaine.

Contre-exemple

Domaine non borné. Considérons la fonction définie par :

$$\forall\, x \in \mathbb{R}, \qquad J(x) = \exp(x)$$

Alors on peut vérifier que

$$\forall\, x \in \mathbb{R}, \qquad J(x) > 0 \qquad \text{et} \qquad \inf_{x\in\mathbb{R}} J(x) = 0$$

Autrement dit, J n'admet pas de minimiseur. Pourtant, cette fonction est convexe et continue sur son domaine, qui est fermé.

La proposition **??** est utile lorsque le domaine de la fonction objectif est borné et qu'elle est continue sur son domaine. Cependant, un grand nombre de problèmes d'optimisation ne vérifie pas cette hypothèse. Dans ce cas, c'est en général la propriété dite de *coercivité* qui sera utilisée. Avant de l'introduire, on établit le résultat général suivant :

Proposition 4

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction continue. Supposons qu'il existe $x^0 \in \mathcal{X}$ tel que l'ensemble de niveau inférieur

$$\operatorname{niv}_{\le J(x^0)} J = \left\{ x \in \mathcal{X} \mid J(x) \le J(x^0) \right\}$$

est borné. Alors J admet un minimiseur.

Démonstration : Par hypothèse, l'ensemble $\operatorname{niv}_{\le J(x^0)} J$ est un ensemble fermé et borné en dimension finie donc un ensemble compact. D'après la proposition précédente, il existe donc $x^* \in \mathcal{X}$ tel que

$$\forall\, x \in \operatorname{niv}_{\le J(x^0)} J, \qquad J(x) \ge J(x^*)$$

En particulier, puisque x^0 appartient à $\operatorname{niv}_{\le J(x^0)} J$, on a

$$J(x^0) \ge J(x^*)$$

Or, par définition de l'ensemble $\operatorname{niv}_{\le J(x^0)} J$, on a

$$\forall\, x \notin \operatorname{niv}_{\le J(x^0)} J, \qquad J(x) > J(x^0) > J(x^*)$$

ce qui prouve que x^* est un minimiseur global de J. ■

Parmi les fonctions qui satisfont les hypothèses de la proposition précédente, il y a les fonctions *coercives* et continues sur leur domaine.

Définition 9 (Fonction coercive)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. On dit que J est *coercive*, ou encore *infinie à l'infini*, si pour toute suite $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$,

$$\lim_{k\to+\infty} \|x_k\| = +\infty \qquad \Longrightarrow \qquad \lim_{k\to+\infty} J(x_k) = +\infty$$

Remarque : Toute fonction $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ de domaine borné est coercive.

Exemple

Fonctions coercives. D'après la définition, la norme euclidienne est une fonction coercive. De plus, si $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ est une fonction telle que

$$\lim_{t\to+\infty} f(t) = +\infty$$

alors $x \mapsto f(\|x\|)$ est une fonction coercive. De manière générale, si J est coercive, alors $f \circ J$ l'est également.

Exemple

Fonctions fortement convexes. Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction fortement convexe de module $\alpha > 0$. Alors J est coercive. En effet, il existe une fonction convexe $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ telle que

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

Or, il est possible de montrer que toute fonction convexe admet une *minorante affine*, c'est-à-dire qu'il existe une fonction affine $g_0 : \mathcal{X} \to \mathbb{R}$ telle que $g \ge g_0$; ainsi, on en déduit que

$$\forall\, x \in \mathcal{X}, \qquad J(x) \ge g_0(x) + \frac{\alpha}{2}\,\|x\|^2 = \sum_{i=1}^{n} \left(a_i\, x_{(i)} + b_i + \frac{\alpha}{2}\, x_{(i)}^2 \right)$$

Ainsi, en notant que $\|x_k\| \to +\infty$ implique que la suite d'au moins une des composantes $((x_{(i)})_k)_{k\in\mathbb{N}}$ diverge vers l'infini, on en déduit le résultat souhaité.

Proposition 5 (Existence de minimiseurs II)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction coercive et continue sur son domaine qui est supposé fermé et non vide. Alors J admet un minimiseur.

Démonstration : Il suffit de montrer que, dans ce cas, la fonction J est telle qu'il existe $x^0 \in \mathcal{X}$ tel que l'ensemble de niveau inférieur

$$\operatorname{niv}_{\le J(x^0)} J = \left\{ x \in \mathcal{X} \mid J(x) \le J(x^0) \right\}$$

soit compact, c'est-à-dire fermé et borné.

- **$\operatorname{niv}_{\le J(x^0)} J$ est fermé.** On commence par noter que $\operatorname{niv}_{\le J(x^0)} J \subset \operatorname{dom} J$. Soit $(x_k)_{k\in\mathbb{N}}$ une suite de $\operatorname{niv}_{\le J(x^0)} J$. On suppose qu'elle converge vers x. Le domaine de J étant fermé, on en déduit que $J(x) \in \operatorname{dom} J$. Ainsi, on a par définition de la suite des x_k et par continuité de J sur son domaine

$$\forall\, k \in \mathbb{N}, \quad J(x_k) \le J(x_0) \qquad \text{et} \qquad \lim_{k\to+\infty} J(x_k) = J(x)$$

  Il suffit de passer à la limite dans l'inégalité pour obtenir que $J(x) \le J(x^0)$.
- **$\operatorname{niv}_{\le J(x^0)} J$ est borné.** On l'établit par l'absurde, en supposant que pour tout $x^0 \in \mathcal{X}$, l'ensemble $\operatorname{niv}_{\le J(x^0)} J$ n'est pas borné. Soit $x^0 \in \mathcal{X}$. Par hypothèse, il existe donc une suite $(x_k)_{k\in\mathbb{N}}$ de $\mathcal{X}$ telle que

$$\lim_{k\to+\infty} \|x_k\| = +\infty \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad J(x_k) \le J(x_0)$$

La suite $(J(x_k))_{k\in\mathbb{N}}$ est donc bornée, en particulier, elle ne peut pas diverger vers $+\infty$, ce qui contredit le fait que J soit coercive.

Ainsi, la fonction J est continue sur le compact $\mathrm{niv}_{\leq J(x^0)} J$, donc elle y admet un minimiseur, noté x^*. Par définition de x^*, on a

$$\forall\, x \notin \mathrm{niv}_{\leq J(x^0)} J, \qquad J(x) > J(x^0) \geq J(x^*)$$

Aussi, x^* est un minimiseur de J sur $\mathcal{X}$. ■

En réalité, il est possible de remplacer l'hypothèse de continuité par celle de semi-continuité inférieure (module A3) :

Proposition 6 (Existence de minimiseurs III)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction coercive et s.c.i. On suppose que son domaine est non vide. Alors J admet un minimiseur.

Démonstration : Soit $(x_k)_{k\in\mathbb{N}} \in (\mathrm{dom}\, J)^{\mathbb{N}}$ une suite minimisante de J (qui existe d'après la proposition **??**). Si cette suite n'est pas bornée, alors la suite des $\|x_k\|$ diverge vers $+\infty$. La coercivité de J assurerait donc que $(J(x_k))_{k\in\mathbb{N}}$ diverge vers $+\infty$, ce qui contredit le fait que cette suite a pour limite la borne inférieure de J (qui appartient à $\mathbb{R} \cup \{-\infty\}$ car $\mathrm{dom}\, J \neq \emptyset$). Par conséquent, la suite $(x_k)_{k\in\mathbb{N}}$ est bornée. Elle admet donc une sous-suite $(x_{k_j})_{j\in\mathbb{N}}$ convergente. Notons x^* sa limite. Puisque J est s.c.i., il s'ensuit que

$$\inf_{x\in\mathcal{X}} J(x) \leq J(x^*) \leq \liminf_{j\to+\infty} J(x_{k_j}) = \lim_{j\to+\infty} J(x_{k_j}) = \inf_{x\in\mathcal{X}} J(x)$$

Autrement dit, x^* est un minimiseur de J. ■

Terminons ce paragraphe avec deux résultats spécifiques aux fonctions convexes.

Proposition 7 (Unicité du minimiseur)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **strictement convexe**. Alors J admet au plus un minimiseur.

Démonstration : Supposons que J admet deux minimiseurs $x_1 \neq x_2$. Ces deux points appartiennent nécessairement au domaine de J. La fonction J étant strictement convexe, on a alors

$$J\left(\frac{x_1 + x_2}{2}\right) < \frac{J(x_1) + J(x_2)}{2} = J(x_1)$$

où le point $(x_1 + x_2)/2$ appartient au domaine, convexe, de J. Ceci est absurde car x_1 est un minimiseur de J. ■

Proposition 8 (Minimiseur d'une fonction fortement convexe)

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. Alors J admet un unique minimiseur.

Démonstration : L'unicité du minimiseur est une conséquence directe de la stricte convexité de J. Il suffit alors de montrer que la fonction J est coercive. Soit $x^0 \in \mathrm{dom}\, f$ tel que $\partial J(x^0)$ soit non vide. Soit $p \in \partial J(x^0)$. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad J(x) \geq J(x^0) + \langle p, x - x^0 \rangle + \frac{\alpha}{2} \|x - x^0\|^2$$

On en déduit par comparaison la coercivité de J en faisant tendre $\|x\|$ vers $+\infty$. ■

2 Condition d'optimalité du premier ordre

2.1 Points critiques

Définition 10 (Point critique ou stationnaire)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. On appelle *point critique* ou *point stationnaire* de J tout point $x^* \in \mathcal{X}$ tel que

$$0 \in \partial J(x^*)$$

On note $\mathrm{crit} J$ l'ensemble des points critiques de J.

Dans le cas différentiable, les points critiques sont exactement les zéros du gradient. Dans le cas d'une fonction réelle continûment dérivable, il existe quatre types de points critiques isolés :

- **les minima locaux** : la dérivée y change de signe, passant du signe négatif au signe positif ;
- **les maxima locaux** : la dérivée y change de signe, passant du signe positif au signe négatif ;
- **les points d'inflexion montants** : la dérivée reste positive au voisinage du point ;
- **les points d'inflexion descendants** : la dérivée reste négative au voisinage du point.

Il existe évidemment des exemples de points critiques non isolés, pour lesquels la dérivée reste nulle sur un voisinage.

2.2 Règle de Fermat

Proposition 9 (Condition nécessaire et suffisante d'optimalité du premier ordre)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**. Soit $x^* \in \mathrm{dom}\, J$. Alors on a l'équivalence entre les deux affirmations suivantes :

(i) x^* est un minimiseur de J ;

(ii) $0 \in \partial J(x^*)$.

Cette condition est également connue sous le nom de *règle / théorème de* Fermat ; on parle parfois d'*équation d'*Euler–Lagrange.

DÉMONSTRATION : Par définition, si x^* est un minimiseur de J si et seulement si

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

c'est-à-dire si et seulement si $0 \in \partial J(x^*)$. ■

Si J n'est pas convexe, alors cette condition n'est plus suffisante :

Proposition 10 (Condition nécessaire d'optimalité du premier ordre)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $x^* \in \operatorname{dom} J$ un minimiseur de J. Alors $0 \in \partial J(x^*)$.

Autrement dit, tout minimiseur est un point critique. La réciproque est en générale fausse, puisqu'il existe des points critiques qui ne sont pas des minimiseurs. Dans le cas différentiable, on peut citer par exemple les minimiseurs locaux ou les maximiseurs (locaux et globaux).

DÉMONSTRATION : En reprenant les calculs de la preuve précédente, on prouve que si x^* est un minimiseur de J, alors $0 \in \hat{\partial} J(x^*) \subset \partial J(x^*)$. ■

3 Algorithmes du premier ordre

3.1 Algorithmes d'optimisation

On formalise ici la notion d'algorithme utilisée dans ce cours.

Définition 11 (Algorithme)

On définit un *algorithme* comme une application $\mathcal{A}$ qui pour tout $k \in \mathbb{N}$, associe au point courant x_k et un ensemble de *paramètres* $\mathcal{P}_k$ (appartenant à un espace vectoriel de dimension finie) un nouveau point x_{k+1}. La suite $(x_k)_{k\in\mathbb{N}} \subset \mathcal{X}^{\mathbb{N}}$ définie de manière récursive par

$$\forall\, k \in \mathbb{N}, \qquad x_{k+1} = \mathcal{A}(x_k; \mathcal{P})$$

est alors dite générée par $\mathcal{A}$. Le point initial x_0 est un des éléments de $\mathcal{P}$.

REMARQUE : Dans cette définition, on suppose que la suite générée par $\mathcal{A}$ est infinie. En pratique, cette suite sera toujours finie.

Un algorithme d'optimisation $\mathcal{A}$ vise à résoudre le problème $(\mathcal{P})$ en générant une suite $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ qui doit, dans le cas idéal, converger vers un point x^* minimiseur de J :

$$\lim_{k\to+\infty} x_k = x^* \qquad \text{avec} \qquad x^* \in \operatorname*{argmin}_{x\in\mathcal{X}} J(x)$$

Un algorithme d'optimisation $\mathcal{A}$ **du premier ordre** quant à lui à trouver un **minimiseur** de J en recherchant un **point critique** de J. Il doit donc être conçu de sorte de générer une suite $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ telle que

$$\lim_{k\to+\infty} x_k = x^* \qquad \text{avec} \qquad 0 \in \partial J(x^*)$$

De manière générale, il faut donc s'assurer que le point critique trouvé est bien un minimiseur de J.

3.2 Modes de convergence

Démontrer la convergence d'un algorithme, c'est démontrer que la suite générée par celui-ci converge bien vers un point satisfaisant les conditions désirées. Dans le cadre des méthodes d'optimisation du premier ordre, ces conditions peuvent être écrites sous les trois formes suivantes

$$\text{(minimiseur)} \qquad x^* \in \operatorname*{argmin}_{x\in\mathcal{X}} J(x) \tag{C1}$$

$$\text{(minimimum)} \qquad J(x^*) = \min_{x\in\mathcal{X}} J(x) \tag{C2}$$

$$\text{(point critique)} \qquad 0 \in \partial J(x^*) \tag{C3}$$

La propriété la plus forte est la première, dans le sens où si x^* est un minimiseur de J, alors $J(x^*)$ est le minimum de J et x^* est un point critique de J. La seconde condition est moins exploitable, car elle ne concerne que la valeur du minimum ; or celle-ci n'est pas suffisante pour déterminer un minimum. Enfin, la dernière condition est la plus faible, car un point critique n'est pas nécessairement un minimiseur. Dans le cas d'une fonction convexe, c'est le cas, mais, de même que connaître la valeur d'un point ne permet pas de le déterminer, savoir qu'un point admet pour (sous-)gradient 0 ne permet pas de le déterminer.

Les preuves de convergence d'algorithmes du premier ordre s'attachent donc à démontrer tout ou partie des propriétés suivantes sur les suites $(x_k)_{k\in\mathbb{N}}$ générées par un algorithme :

- convergence de la suite des itérés x_k :

$$\lim_{k\to+\infty} x_k = x^* \tag{A-I}$$

- convergence des itérés vers un minimiseur de J :

$$\lim_{k\to+\infty} x_k = x^* \qquad \text{avec} \qquad x^* \in \operatorname*{argmin}_{x\in\mathcal{X}} J(x) \tag{A-II}$$

- convergence des itérés vers un point critique de J :

$$\lim_{k\to+\infty} x_k = x^* \qquad \text{avec} \qquad 0 \in \partial J(x^*) \tag{A-III}$$

- convergence en valeur / du critère :

$$\lim_{k\to+\infty} J(x_k) = J^* \tag{B-I}$$

- convergence en valeur / du critère vers le minimum de J

$$\lim_{k\to+\infty} J(x_k) = \min_{x\in\mathcal{X}} J(x) \tag{B-II}$$

- convergence vers le critère d'optimalité / des sous-gradients :

$$\forall\, k \in \mathbb{N}, \exists\, p_k \in \partial J(x_k), \qquad \lim_{k\to+\infty} p_k = p^* \tag{C-I}$$

- convergence vers le critère d'optimalité / des sous-gradients vers 0 :

$$\forall\, k \in \mathbb{N}, \exists\, p_k \in \partial J(x_k), \qquad \lim_{k\to+\infty} p_k = 0 \tag{C-II}$$

Notons qu'il n'est pas toujours possible de démontrer tous ces modes de convergence pour un même algorithme. Toutefois, sous certaines hypothèses sur le problème (sur la fonction objectif J), il existe des liens naturels entre certains modes de convergence. On en cite quelques uns ici pour exemple :

Proposition 11 (Cas continu)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $(x_k)_{k\in\mathbb{N}}$ une suite générée par un algorithme $\mathcal{A}$. Soit $x^* \in \operatorname{dom} J$. On suppose que l'une des deux hypothèses suivantes est vérifiée :

(i) le point $x^* \in \operatorname{int}(\operatorname{dom} J)$ et J est continue au voisinage de x^* ;

(ii) la fonction J est continue sur son domaine et il existe $k_0 \in \mathbb{N}$ tel que

$$\forall\, k \geq k_0, \qquad x_k \in \operatorname{dom} J$$

Alors
$$\lim_{k\to+\infty} x_k = x^* \qquad \Longrightarrow \qquad \lim_{k\to+\infty} J(x_k) = J(x^*)$$

De plus, si $x^* \in \operatorname{int}(\operatorname{dom} J)$ est un minimiseur de J, alors $J(x^*)$ est le minimum de J.

Autrement dit, la convergence des itérées (A-I) (resp. vers un minimiseur (A-II)) implique la convergence en valeur (B-I) (resp. vers le minimum (B-II)). De la même façon, on a dans le cas d'une fonction continûment différentiable :

Proposition 12 (Cas continûment différentiable)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $(x_k)_{k\in\mathbb{N}}$ une suite générée par un algorithme $\mathcal{A}$. On suppose que J est continûment différentiable au voisinage de x^*. Alors

$$\lim_{k\to+\infty} x_k = x^* \qquad \Longrightarrow \qquad \lim_{k\to+\infty} \nabla J(x_k) = \nabla J(x^*)$$

De plus, si $x^* \in \operatorname{int}(\operatorname{dom} J)$ est un minimiseur de J ou un point critique de J, alors $\nabla J(x^*) = 0$.

Autrement dit, la convergence des itérées (A-I) (resp. vers un minimiseur (A-II) ou vers un point critique (A-III)) implique la convergence des gradients (C-I) (resp. vers 0 (C-II)). Dans le cas convexe, on a le résultat suivant :

Proposition 13 (Cas convexe)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** propre. On suppose que J admet un minimiseur. Soit $(x_k)_{k\in\mathbb{N}}$ une suite générée par un algorithme $\mathcal{A}$. On suppose que la suite $(x_k)_{k\in\mathbb{N}}$ est bornée et que

$$\forall\, k \in \mathbb{N}, \qquad p_k \in \partial J(x_k) \qquad \text{avec} \qquad \lim_{k\to+\infty} p_k = 0$$

Alors la suite $(J(x_k))_{k\in\mathbb{N}}$ converge vers le minimum de J.

Démonstration : Soit $k \in \mathbb{N}$ et x^* un minimiseur de J. La convexité de J assure que
$$J(x^*) \geq J(x_k) + \langle p_k, x^* - x_k\rangle$$
L'inégalité de Cauchy–Schwarz permet d'obtenir la majoration suivante :
$$0 \leq J(x_k) - J(x^*) \leq \|\nabla J(x_k)\|\, \|x_k - x^*\|$$
Par comparaison, on obtient le résultat désiré. ■

Ce dernier résultat nécessite de montrer que la suite générée par l'algorithme est bornée. Ce n'est pas toujours le cas, ni évident à établir. Toutefois, on verra de nombreux cas où cette propriété est vérifiée. On peut d'ores-et-déjà citer le cas où les points x_k appartiennent au domaine de J lorsque celui-ci est borné.

Grâce à la fermeture du sous-différentiel, on peut démontrer le résultat suivant :

Proposition 14

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $(x_k)_{k\in\mathbb{N}}$ une suite générée par un algorithme $\mathcal{A}$. Soit $x^* \in \operatorname{dom} J$. On suppose que

$$\lim_{k\to+\infty} x_k = x^* \qquad \text{et} \qquad \lim_{k\to+\infty} J(x_k) = J(x^*)$$

et que
$$\forall\, k \in \mathbb{N}, \qquad p_k \in \partial J(x_k) \qquad \text{avec} \qquad \lim_{k\to+\infty} p_k = 0$$

Alors
$$0 \in \partial J(x^*)$$

Autrement dit, la convergence des itérées (A-I), du critère d'optimalité (C-I) et des sous-gradients vers 0 (B-II) implique la convergence des itérés vers un point critique (A-III).

Intuitivement, il apparaît que la convergence des itérées peut être considérée comme la plus forte, car, dans les applications, les fonctions sont souvent au moins continues, voir continûment différentiable. Cependant, il s'agit généralement du mode de convergence le plus difficile à établir.

Les différents modes de convergence permettent de définir des *critères d'arrêt* pour les algorithmes, c'est-à-dire de définir une condition qui, une fois satisfaite, entraîne l'arrêt des itérations. Un bon critère d'arrêt doit stopper les itérations en un temps fini raisonnable (quelques secondes, minutes, heures en général, suivant les applications), aboutir à une itération suffisamment proche du point recherché et ne doit pas être trop complexe à calculer.

Ainsi par exemple, si on a démontré que la convergence des sous-gradients vers 0 (C-II), un critère d'arrêt pourrait être défini à l'aide d'un seuil dit de *tolérance* $\varepsilon > 0$, qui permettrait d'arrêter les itérations dès que la condition suivante est satisfaite pour un certain $k \in \mathbb{N}$:
$$\|p_k\| < \varepsilon \qquad \text{avec} \qquad p_k \in \partial J(x_k)$$
Ce critère est exploitable à condition qu'un tel p_k soit calculable (et on a vu dans le module A2 qu'il est parfois difficile à déterminer) et avec une complexité raisonnable. On verra dans les modules qui suivent que, pour de nombreux algorithmes, c'est le cas. Évidemment, si ε est choisi trop grand, alors le point final obtenu x_k peut être trop éloigné d'un minimiseur (ou d'un point critique) de J. En revanche, si ε est trop petit, il faudra peut-être trop d'itérations pour satisfaire le critère d'arrêt défini. En pratique,

le choix de la valeur ε est parfois difficile, d'autant que le lien entre la valeur $\|p_k\|$ et la distance de x_k au minimiseur / point critique x^* (la cas échéant) n'est pas toujours explicite.

Pour s'affranchir de cette difficulté, on serait tenté d'utiliser un autre mode de convergence. Le plus naturel est d'utiliser la convergence des itérés x_k vers un point x^* (A). En effet, définir comme critère d'arrêt

$$\|x_k - x^*\| < \varepsilon$$

permet de contrôler très précisément l'erreur commise. Malheureusement, un tel critère d'arrêt nécessite de connaître le point x^*, ce qui n'est évidemment pas le cas puisque c'est justement le point recherché. Un moyen simple d'utiliser la convergence des itérés (si elle est démontrée) est de définir plutôt un critère d'arrêt de la forme

$$\|x_{k+1} - x_k\| < \varepsilon$$

Malheureusement, il est alors difficile d'en déduire l'erreur commise à l'issu des itérations. Si ε est choisi très petit (par exemple $\varepsilon = 10^{-9}$), on peut numériquement estimer que le point limite x^* est atteint, car la suite des itérés reste constante à un arrondi près. Les mêmes remarques sont valables si on utilise la convergence du critère (B).

Notons enfin qu'un bon moyen de majorer le temps de calcul est d'inclure dans le critère d'arrêt une autre condition sur le nombre d'itérations total, de sorte que l'algorithme s'arrête si la tolérance ε est atteinte ou si le nombre d'itérations maximal est atteint.

3.3 Taux de convergence

Pour terminer, on introduit une notion importante sur la vitesse de convergence d'un algorithme. On prend l'exemple de la convergence des itérés.

Définition 12 (Taux de convergence)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $(x_k)_{k\in\mathbb{N}}$ une suite générée par un algorithme $\mathcal{A}$. On suppose qu'elle converge vers x^*, et qu'elle ne converge pas en un nombre fini d'itérations. La convergence de l'algorithme $\mathcal{A}$ est dite :

- *linéaire* s'il existe $\tau \in \,]\,0\,;1\,[$ tel que

$$\lim_{k\to+\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \tau$$

- *d'ordre* p s'il existe $\tau \in [\,0\,;+\infty\,[$ tel que

$$\lim_{k\to+\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} = \tau$$

Remarque : Si $p = 2$, on parle de taux de convergence *quadratique*.

La convergence linéaire est la plus intéressante, mais généralement réservée à des problèmes très réguliers. Elle se traduit par une décroissante exponentielle de l'erreur (ici $\|x_k - x^*\|$). On peut parfois démontrer des taux de convergence linéaire (ou d'ordre p) pour d'autres quantités.

Pour aller plus loin

Optimisation sous contraintes. L'introduction de la droite achevée permet de transformer les problèmes d'optimisation sous contraintes en problèmes non contraints. Plus précisément, elle permet d'utiliser un formalisme commun pour étudier les deux types de problèmes, en incorporant dans la fonction objectif les contraintes à l'aide de l'indicatrice de l'ensemble admissible. De fait, un problème sous contraintes (différentiable ou non) peut donc toujours être vu comme un problème d'optimisation non lisse non contraint.

Modes de convergence. Pour chacune des méthodes d'optimisation que l'on étudiera dans les modules suivront, on s'attachera à démontrer les différents modes de convergence présentés dans ce module.