Master 1 MAPI3 - IMA - RO

Travaux dirigés d'optimisation

PIERRE MARÉCHAL

Université Paul Sabatier

pr.marechal@gmail.com

Existence et conditions d’optimalité

Algorithmes

Exercice

Soient $r \in \mathbb{R}_+^*$, A une matrice réelle symétrique définie positive de taille $n \times n$ et

$$K := \{\mathbf{x} \in \mathbb{R}^n \,|\, \langle \mathbf{x}, A\mathbf{x} \rangle \leq 2r\}.$$

Soit $f : \mathbb{R}^n \to \mathbb{R}$ définie par $f(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x} \rangle$ avec $\mathbf{b} \in \mathbb{R}^n$.

(1) Montrer qu'il existe $\lambda_{\min} > 0$ et $\lambda_{\max} > 0$ tels que

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad \lambda_{\min} \|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle \leq \lambda_{\max} \|\mathbf{x}\|^2.$$

(2) Montrer que K est compact.

(3) Montrer que f atteint son minimum sur K.

(1) Rappelons que toute matrice réelle symétrique est diagonalisable dans une base orthonormale, et que si la matrice est définie positive, ses valeurs propres sont strictement positives.

(1) Rappelons que toute matrice réelle symétrique est diagonalisable dans une base orthonormale, et que si la matrice est définie positive, ses valeurs propres sont strictement positives. Il existe donc $(\lambda_1, \mathbf{v}_1), \ldots, (\lambda_n, \mathbf{v}_n) \in \mathbb{R}_+^* \times \mathbb{R}^n$ tels que

   (i) $\lambda_1 \geq \ldots \geq \lambda_n > 0$;
   (ii) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ est une base orthonormale de $\mathbb{R}^n$;
   (iii) pour tout $k \in \{1, \ldots, n\}$, $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$.

Soit alors $\mathbf{x}$ quelconque dans $\mathbb{R}^n$.

Soit alors $\mathbf{x}$ quelconque dans $\mathbb{R}^n$. Soient $\alpha_1, \ldots, \alpha_n$ les composantes de $\mathbf{x}$ dans la base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Soit alors $\mathbf{x}$ quelconque dans $\mathbb{R}^n$. Soient $\alpha_1, \ldots, \alpha_n$ les composantes de $\mathbf{x}$ dans la base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. On vérifie facilement que

$$\|\mathbf{x}\|^2 = \sum_{k=1}^{n} \alpha_k^2 \quad \text{et} \quad \langle \mathbf{x}, A\mathbf{x} \rangle = \sum_{k=1}^{n} \lambda_k \alpha_k^2.$$

Soit alors $\mathbf{x}$ quelconque dans $\mathbb{R}^n$. Soient $\alpha_1, \ldots, \alpha_n$ les composantes de $\mathbf{x}$ dans la base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. On vérifie facilement que

$$\|\mathbf{x}\|^2 = \sum_{k=1}^{n} \alpha_k^2 \quad \text{et} \quad \langle \mathbf{x}, A\mathbf{x} \rangle = \sum_{k=1}^{n} \lambda_k \alpha_k^2.$$

Pour tout $k \in \{1, \ldots, n\}$, $\lambda_n \alpha_k^2 \leq \lambda_k \alpha_k^2 \leq \lambda_1 \alpha_k^2$.

Soit alors $\mathbf{x}$ quelconque dans $\mathbb{R}^n$. Soient $\alpha_1, \ldots, \alpha_n$ les composantes de $\mathbf{x}$ dans la base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. On vérifie facilement que

$$\|\mathbf{x}\|^2 = \sum_{k=1}^{n} \alpha_k^2 \quad \text{et} \quad \langle \mathbf{x}, A\mathbf{x} \rangle = \sum_{k=1}^{n} \lambda_k \alpha_k^2 .$$

Pour tout $k \in \{1, \ldots, n\}$, $\lambda_n \alpha_k^2 \leq \lambda_k \alpha_k^2 \leq \lambda_1 \alpha_k^2$. En sommant ces dernières inégalités, nous obtenons:

$$\lambda_n \|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle \leq \lambda_1 \|\mathbf{x}\|^2 .$$

L'inégalité demandée est donc satisfaite pour $\lambda_{\min} = \lambda_n$ et $\lambda_{\max} = \lambda_1$.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle \leq 2r$.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0, 2r])$, avec

$$\varphi : \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle. \end{array}$$

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0, 2r])$, avec

$$\begin{array}{rccc} \varphi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle . \end{array}$$

La fonction φ est continue.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0, 2r])$, avec

$$\begin{array}{rccc} \varphi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle . \end{array}$$

La fonction φ est continue. En effet, pour tous $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$,

$$\begin{aligned} |\varphi(\mathbf{x}+\mathbf{h}) - \varphi(\mathbf{x})| &= |2\langle A\mathbf{x}, \mathbf{h}\rangle + \langle \mathbf{h}, A\mathbf{h}\rangle| \\ &\leq 2\|A\mathbf{x}\| \cdot \|\mathbf{h}\| + \lambda_{\max}\|\mathbf{h}\|^2, \end{aligned}$$

ce qui montre que $\varphi(\mathbf{x}+\mathbf{h}) \to \varphi(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0,2r])$, avec

$$\begin{array}{rccc} \varphi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle . \end{array}$$

La fonction φ est continue. En effet, pour tous $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$,

$$\begin{aligned} |\varphi(\mathbf{x}+\mathbf{h}) - \varphi(\mathbf{x})| &= |2\langle A\mathbf{x}, \mathbf{h}\rangle + \langle \mathbf{h}, A\mathbf{h}\rangle| \\ &\leq 2\|A\mathbf{x}\| \cdot \|\mathbf{h}\| + \lambda_{\max}\|\mathbf{h}\|^2, \end{aligned}$$

ce qui montre que $\varphi(\mathbf{x}+\mathbf{h}) \to \varphi(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$. L'ensemble K est donc l'image réciproque de l'intervalle fermé $[0,2r]$ par l'application continue φ.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0,2r])$, avec

$$\begin{array}{cccc} \varphi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle . \end{array}$$

La fonction φ est continue. En effet, pour tous $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$,

$$\begin{aligned} |\varphi(\mathbf{x}+\mathbf{h}) - \varphi(\mathbf{x})| &= |2\langle A\mathbf{x}, \mathbf{h}\rangle + \langle \mathbf{h}, A\mathbf{h}\rangle| \\ &\leq 2\|A\mathbf{x}\| \cdot \|\mathbf{h}\| + \lambda_{\max}\|\mathbf{h}\|^2, \end{aligned}$$

ce qui montre que $\varphi(\mathbf{x}+\mathbf{h}) \to \varphi(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$. L'ensemble K est donc l'image réciproque de l'intervalle fermé $[0,2r]$ par l'application continue φ. C'est donc un fermé de $\mathbb{R}^n$.

(2) Pour tout $\mathbf{x} \in K$, $\lambda_{\min}\|\mathbf{x}\|^2 \leq \langle \mathbf{x}, A\mathbf{x}\rangle \leq 2r$. On en déduit que K est contenu dans la boule de centre $\mathbf{0}$ et de rayon $\sqrt{2r/\lambda_{\min}}$. Donc K est borné. D'autre part, $K = \varphi^{-1}([0, 2r])$, avec

$$\begin{array}{rccc} \varphi : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \longmapsto & \langle \mathbf{x}, A\mathbf{x}\rangle. \end{array}$$

La fonction φ est continue. En effet, pour tous $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$,

$$\begin{aligned} |\varphi(\mathbf{x}+\mathbf{h}) - \varphi(\mathbf{x})| &= |2\langle A\mathbf{x}, \mathbf{h}\rangle + \langle \mathbf{h}, A\mathbf{h}\rangle| \\ &\leq 2\|A\mathbf{x}\| \cdot \|\mathbf{h}\| + \lambda_{\max}\|\mathbf{h}\|^2, \end{aligned}$$

ce qui montre que $\varphi(\mathbf{x}+\mathbf{h}) \to \varphi(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$. L'ensemble K est donc l'image réciproque de l'intervalle fermé $[0, 2r]$ par l'application continue φ. C'est donc un fermé de $\mathbb{R}^n$. Puisque K est aussi borné, c'est un compact.

(3) L'application f est continue.

(3) L'application f est continue. En effet,

$$\begin{aligned} |f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})| &= |\langle \mathbf{b}, \mathbf{h} \rangle| \\ &\leq \|\mathbf{b}\| \cdot \|\mathbf{h}\|, \end{aligned}$$

ce qui montre que $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$.

(3) L'application f est continue. En effet,

$$\begin{aligned} |f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})| &= |\langle \mathbf{b},\mathbf{h}\rangle| \\ &\leq \|\mathbf{b}\|\cdot\|\mathbf{h}\|, \end{aligned}$$

ce qui montre que $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$. On en déduit que f atteint son minimum sur tout compact, en particulier sur K.

Exercice

Soient A une matrice réelle symétrique définie positive de taille $n \times n$, $\mathbf{b} \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Soit $f : \mathbb{R}^n \to \mathbb{R}$ l'application définie par

$$f(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x} \rangle - \langle \mathbf{b}, \mathbf{x} \rangle + c.$$

(1) Montrer que f est continue.

(2) Montrer que, pour tout $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) \geq \frac{1}{2}\lambda_{\min}\|\mathbf{x}\|^2 - \|\mathbf{b}\| \cdot \|\mathbf{x}\| + c.$$

(3) Montrer que f atteint son minimum sur $\mathbb{R}^n$.

(1) On a:

$$f(\mathbf{x}+\mathbf{h}) = \frac{1}{2}\left[\langle A\mathbf{x},\mathbf{x}\rangle + 2\langle A\mathbf{x},\mathbf{h}\rangle + \langle A\mathbf{h},\mathbf{h}\rangle\right] - \langle \mathbf{b},\mathbf{x}\rangle - \langle \mathbf{b},\mathbf{h}\rangle + c.$$

(1) On a:

$$f(\mathbf{x}+\mathbf{h}) = \frac{1}{2}\left[\langle A\mathbf{x},\mathbf{x}\rangle + 2\langle A\mathbf{x},\mathbf{h}\rangle + \langle A\mathbf{h},\mathbf{h}\rangle\right] - \langle \mathbf{b},\mathbf{x}\rangle - \langle \mathbf{b},\mathbf{h}\rangle + c.$$

Donc

$$\begin{aligned} |f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})| &= \left|\langle A\mathbf{x}-\mathbf{b},\mathbf{h}\rangle + \frac{1}{2}\langle A\mathbf{h},\mathbf{h}\rangle\right| \\ &\leq \|A\mathbf{x}-\mathbf{b}\|\cdot\|\mathbf{h}\| + \frac{\lambda_{\max}}{2}\|\mathbf{h}\|^2. \end{aligned}$$

(1) On a:

$$f(\mathbf{x}+\mathbf{h}) = \frac{1}{2}\left[\langle A\mathbf{x},\mathbf{x}\rangle + 2\langle A\mathbf{x},\mathbf{h}\rangle + \langle A\mathbf{h},\mathbf{h}\rangle\right] - \langle \mathbf{b},\mathbf{x}\rangle - \langle \mathbf{b},\mathbf{h}\rangle + c.$$

Donc

$$\begin{aligned} |f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})| &= \left|\langle A\mathbf{x}-\mathbf{b},\mathbf{h}\rangle + \frac{1}{2}\langle A\mathbf{h},\mathbf{h}\rangle\right| \\ &\leq \|A\mathbf{x}-\mathbf{b}\|\cdot\|\mathbf{h}\| + \frac{\lambda_{\max}}{2}\|\mathbf{h}\|^2. \end{aligned}$$

On voit donc que $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$ lorsque $\mathbf{h} \to \mathbf{0}$.

(2) L'inégalité demandée résulte immédiatement des inégalités

$$\langle \mathbf{x}, A\mathbf{x} \rangle \geq \lambda_{\min} \|\mathbf{x}\|^2 \quad \text{et} \quad \langle \mathbf{b}, \mathbf{x} \rangle \leq \|\mathbf{b}\| \cdot \|\mathbf{x}\|.$$

(2) L'inégalité demandée résulte immédiatement des inégalités

$$\langle \mathbf{x}, A\mathbf{x} \rangle \geq \lambda_{\min} \|\mathbf{x}\|^2 \quad \text{et} \quad \langle \mathbf{b}, \mathbf{x} \rangle \leq \|\mathbf{b}\| \cdot \|\mathbf{x}\|.$$

(3) D'après la question précédente,

$$\frac{f(\mathbf{x})}{\|\mathbf{x}\|} \geq \frac{1}{2}\lambda_{\min}\|\mathbf{x}\| - \|\mathbf{b}\| + \frac{c}{\|\mathbf{x}\|} \to \infty \quad \text{lorsque} \quad \|\mathbf{x}\| \to \infty.$$

(2) L'inégalité demandée résulte immédiatement des inégalités

$$\langle \mathbf{x}, A\mathbf{x} \rangle \geq \lambda_{\min} \|\mathbf{x}\|^2 \quad \text{et} \quad \langle \mathbf{b}, \mathbf{x} \rangle \leq \|\mathbf{b}\| \cdot \|\mathbf{x}\|.$$

(3) D'après la question précédente,

$$\frac{f(\mathbf{x})}{\|\mathbf{x}\|} \geq \frac{1}{2}\lambda_{\min}\|\mathbf{x}\| - \|\mathbf{b}\| + \frac{c}{\|\mathbf{x}\|} \to \infty \quad \text{lorsque} \quad \|\mathbf{x}\| \to \infty.$$

Autrement dit, f est 1-coercive (donc coercive).

(2) L'inégalité demandée résulte immédiatement des inégalités

$$\langle \mathbf{x}, A\mathbf{x} \rangle \geq \lambda_{\min} \|\mathbf{x}\|^2 \quad \text{et} \quad \langle \mathbf{b}, \mathbf{x} \rangle \leq \|\mathbf{b}\| \cdot \|\mathbf{x}\|.$$

(3) D'après la question précédente,

$$\frac{f(\mathbf{x})}{\|\mathbf{x}\|} \geq \frac{1}{2}\lambda_{\min}\|\mathbf{x}\| - \|\mathbf{b}\| + \frac{c}{\|\mathbf{x}\|} \to \infty \quad \text{lorsque} \quad \|\mathbf{x}\| \to \infty.$$

Autrement dit, f est 1-coercive (donc coercive). Puisque f est continue sur $\mathbb{R}^n$, f atteint son minimum sur $\mathbb{R}^n$.

Le but de l'exercice suivant est de donner une preuve variationnelle du théorème fondamental de l'algèbre, dont voici l'énoncé:

> *Tout polynôme d'une variable complexe, à coefficients complexes et non constant, admet au moins une racine.*

On rappelle que pour tout $z_0 \in \mathbb{C}$, la famille

$$\{(Z - z_0)^k | k \in \{0, \ldots, m\}\}$$

est une base de $\mathbb{C}_m[Z]$, l'espace des polynômes complexes d'une variable, de degré inférieur ou égal à m.

Exercice

(1) Montrer que, quelque soit $a, b \in \mathbb{C}$, $|a-b| \geq |a| - |b|$.

(2) Soient $m \in \mathbb{N}^*$ et $P = a_0 + a_1 Z + \cdots + a_m Z^m \in \mathbb{C}_m[Z]$ tel que $a_m \neq 0$. Montrer que la fonction $z \mapsto |P(z)|$ est 0-coercive, c'est-à-dire, que $|P(z)| \to \infty$ lorsque $|z| \to \infty$.

(3) Montrer que $P(z)$ peut se mettre sous la forme $P(z) = b_0 + b_1(z-z_0) + \ldots + b_m(z-z_0)^m$, où z_0 est un minimiseur (global) de la fonction $z \mapsto |P(z)|$ et $b_0, \ldots, b_m$ sont des nombres complexes avec $b_m \neq 0$.

(4) On suppose maintenant, en vue d'obtenir une contradiction, que $b_0 \neq 0$. Soit k le plus petit indice dans $\{1, \ldots, m\}$ tel que $b_k \neq 0$. Montrer qu'il existe $r > 0$ tel que, pour tout z sur le cercle $\mathscr{C}(z_0, r) := \{z' \in \mathbb{C} \,|\, |z' - z_0| = r\}$,

$$|P(z) - b_0 - b_k(z-z_0)^k| < |b_k| r^k < |b_0|,$$

et conclure.

(1) C'est une conséquence immédiate de l'inégalité triangulaire, appliquée au nombre complexe $a = a - b + b$.

(1) C'est une conséquence immédiate de l'inégalité triangulaire, appliquée au nombre complexe $a = a - b + b$.

(2) D'après la question précédente et l'inégalité triangulaire, pour tout $z \neq 0$,

$$\begin{aligned}
|P(z)| &\geq |a_m z^m| - \left| -\sum_{j=0}^{m-1} a_j z^j \right| \\
&\geq |a_m z^m| - \sum_{j=0}^{m-1} |a_j z^j| \\
&= |a_m||z|^m \left(1 - \sum_{j=0}^{m-1} \frac{|a_j|}{|a_m|} |z|^{j-m} \right).
\end{aligned}$$

(1) C'est une conséquence immédiate de l'inégalité triangulaire, appliquée au nombre complexe $a = a - b + b$.

(2) D'après la question précédente et l'inégalité triangulaire, pour tout $z \neq 0$,

$$\begin{aligned}
|P(z)| &\geq |a_m z^m| - \left| -\sum_{j=0}^{m-1} a_j z^j \right| \\
&\geq |a_m z^m| - \sum_{j=0}^{m-1} |a_j z^j| \\
&= |a_m||z|^m \left(1 - \sum_{j=0}^{m-1} \frac{|a_j|}{|a_m|} |z|^{j-m} \right).
\end{aligned}$$

Le facteur entre parenthèse tend vers 1 lorsque $|z| \to \infty$. On en déduit que $|P(z)| \to \infty$ lorsque $|z| \to \infty$.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global. Puisque $\{(Z-z_0)^k | k \in \{0,\ldots,m\}\}$ est une base de $\mathbb{C}_m[Z]$, on peut mettre $P(z)$ sous la forme demandée, avec $b_m = a_m \neq 0$

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global. Puisque $\{\,(Z-z_0)^k \,|\, k \in \{0,\ldots,m\}\,\}$ est une base de $\mathbb{C}_m[Z]$, on peut mettre $P(z)$ sous la forme demandée, avec $b_m = a_m \neq 0$

(4) Supposons que $b_0 \neq 0$.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global. Puisque $\{(Z-z_0)^k | k \in \{0,\dots,m\}\}$ est une base de $\mathbb{C}_m[Z]$, on peut mettre $P(z)$ sous la forme demandée, avec $b_m = a_m \neq 0$

(4) Supposons que $b_0 \neq 0$. Si $k = m$, alors $|P(z) - b_0 - b_m(z-z_0)^m| = 0$ et le résultat demandé est évident.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global. Puisque $\{(Z-z_0)^k | k \in \{0,\ldots,m\}\}$ est une base de $\mathbb{C}_m[Z]$, on peut mettre $P(z)$ sous la forme demandée, avec $b_m = a_m \neq 0$

(4) Supposons que $b_0 \neq 0$. Si $k = m$, alors $|P(z) - b_0 - b_m(z-z_0)^m| = 0$ et le résultat demandé est évident. Supposons donc que $k \in \{1,\ldots,m-1\}$.

(3) Puisque la fonction $z \mapsto |P(z)|$ est continue et 0-coercive, elle admet un minimiseur global sur $\mathbb{C}$. Soit z_0 ce minimiseur global. Puisque $\{(Z-z_0)^k | k \in \{0,\dots,m\}\}$ est une base de $\mathbb{C}_m[Z]$, on peut mettre $P(z)$ sous la forme demandée, avec $b_m = a_m \neq 0$

(4) Supposons que $b_0 \neq 0$. Si $k = m$, alors $|P(z) - b_0 - b_m(z-z_0)^m| = 0$ et le résultat demandé est évident. Supposons donc que $k \in \{1,\dots,m-1\}$. Puisque

$$|P(z) - b_0 - b_k(z-z_0)^k| = |b_{k+1}(z-z_0)^{k+1} + \cdots + b_m(z-z_0)^m|,$$

il est clair que l'on peut trouver $r > 0$ suffisamment petit pour que

$$\forall z \in \mathscr{C}(z_0, r), \quad |P(z) - b_0 - b_k(z-z_0)^k| < |b_k| r^k < |b_0|. \quad (1)$$

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z - z_0)^k| + |b_0 + b_k(z - z_0)^k|.$$

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z-z_0)^k| + |b_0 + b_k(z-z_0)^k|.$$

Choisissons $z \in \mathscr{C}(z_0, r)$ tel que b_0 et $b_k(z-z_0)^k$ soient de directions opposées dans le plan complexe.

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z-z_0)^k| + |b_0 + b_k(z-z_0)^k|.$$

Choisissons $z \in \mathscr{C}(z_0, r)$ tel que b_0 et $b_k(z-z_0)^k$ soient de directions opposées dans le plan complexe. En tenant compte de l'inégalité $|b_k|r^k < |b_0|$,

$$|b_0 + b_k(z-z_0)^k| = |b_0| - |b_k(z-z_0)^k|,$$

et la majoration (1) implique alors que

$$|P(z)| < |b_k|r^k + |b_0| - |b_k(z-z_0)^k| = |b_0| = |P(z_0)|.$$

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z - z_0)^k| + |b_0 + b_k(z - z_0)^k|.$$

Choisissons $z \in \mathscr{C}(z_0, r)$ tel que b_0 et $b_k(z - z_0)^k$ soient de directions opposées dans le plan complexe. En tenant compte de l'inégalité $|b_k|r^k < |b_0|$,

$$|b_0 + b_k(z - z_0)^k| = |b_0| - |b_k(z - z_0)^k|,$$

et la majoration (1) implique alors que

$$|P(z)| < |b_k|r^k + |b_0| - |b_k(z - z_0)^k| = |b_0| = |P(z_0)|.$$

Ceci contredit le fait que z_0 est un minimiseur global de $|P(z)|$.

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z-z_0)^k| + |b_0 + b_k(z-z_0)^k|.$$

Choisissons $z \in \mathscr{C}(z_0, r)$ tel que b_0 et $b_k(z-z_0)^k$ soient de directions opposées dans le plan complexe. En tenant compte de l'inégalité $|b_k|r^k < |b_0|$,

$$|b_0 + b_k(z-z_0)^k| = |b_0| - |b_k(z-z_0)^k|,$$

et la majoration (1) implique alors que

$$|P(z)| < |b_k|r^k + |b_0| - |b_k(z-z_0)^k| = |b_0| = |P(z_0)|.$$

Ceci contredit le fait que z_0 est un minimiseur global de $|P(z)|$. On en déduit que $b_0 = 0$, puis que $P(z) = (z-z_0)Q(z)$ avec Q de degré $m-1$.

D'après l'inégalité triangulaire,

$$|P(z)| \leq |P(z) - b_0 - b_k(z-z_0)^k| + |b_0 + b_k(z-z_0)^k|.$$

Choisissons $z \in \mathscr{C}(z_0, r)$ tel que b_0 et $b_k(z-z_0)^k$ soient de directions opposées dans le plan complexe. En tenant compte de l'inégalité $|b_k|r^k < |b_0|$,

$$|b_0 + b_k(z-z_0)^k| = |b_0| - |b_k(z-z_0)^k|,$$

et la majoration (1) implique alors que

$$|P(z)| < |b_k|r^k + |b_0| - |b_k(z-z_0)^k| = |b_0| = |P(z_0)|.$$

Ceci contredit le fait que z_0 est un minimiseur global de $|P(z)|$. On en déduit que $b_0 = 0$, puis que $P(z) = (z-z_0)Q(z)$ avec Q de degré $m-1$. Le théorème fondamental de l'algèbre en découle.

Exercice

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = x^2 + y^2 - 2xy + \operatorname{ch} y$.

(1) Calculer le gradient et la hessienne de f en tout point (x,y) de $\mathbb{R}^2$.

(2) Déterminer l'ensemble des points stationnaires de f.

(3) Pour chaque point stationnaire, dire s'il s'agit d'un minimiseur local de f, d'un minimiseur local strict de f.

(1) On vérifie sans peine que f est de classe $\mathscr{C}^2(\mathbb{R}^2)$ et que, pour tout $(x,y) \in \mathbb{R}^2$,

$$\nabla f(x,y) = \begin{pmatrix} 2x - 2y \\ 2y - 2x + \operatorname{sh} y \end{pmatrix}$$

et

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & -2 \\ -2 & 2 + \operatorname{ch} y \end{bmatrix}.$$

(2) On a:

$$\nabla f(x,y) = 0 \iff \left\{ \begin{array}{lcl} 0 & = & 2x - 2y \\ 0 & = & 2y - 2x + \operatorname{sh} y \end{array} \right.$$

$$\iff \left\{ \begin{array}{lcl} x & = & y \\ 0 & = & \operatorname{sh} y \end{array} \right.$$

$$\iff x = y = 0.$$

(2) On a:

$$\nabla f(x,y) = 0 \iff \left\{ \begin{array}{lcl} 0 & = & 2x - 2y \\ 0 & = & 2y - 2x + \operatorname{sh} y \end{array} \right.$$

$$\iff \left\{ \begin{array}{lcl} x & = & y \\ 0 & = & \operatorname{sh} y \end{array} \right.$$

$$\iff x = y = 0.$$

Donc l'ensemble des points stationnaires de f est $\{(0,0)\}$.

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix}.$$

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix}.$$

Cette matrice est définie positive, puisque sa trace et son déterminant sont strictement positifs (ce qui caractérise, rappelons-le, la définie positivité pour les matrices de taille 2×2).

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix}.$$

Cette matrice est définie positive, puisque sa trace et son déterminant sont strictement positifs (ce qui caractérise, rappelons-le, la définie positivité pour les matrices de taille 2×2). Le théorème du cours sur les conditions suffisantes d'optimalité montre alors que $(0,0)$ est un minimiseur local strict de f.

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix}.$$

Cette matrice est définie positive, puisque sa trace et son déterminant sont strictement positifs (ce qui caractérise, rappelons-le, la définie positivité pour les matrices de taille 2×2). Le théorème du cours sur les conditions suffisantes d'optimalité montre alors que $(0,0)$ est un minimiseur local strict de f.
On remarque que $(0,0)$ est aussi un minimiseur global strict puisque, pour tout $(x,y) \neq (0,0)$,

$$f(x,y) = (x-y)^2 + \operatorname{ch} y > 1 = f(0,0).$$

Exercice

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = (x+y)^3 + \mathrm{ch}(x-y)$.

(1) Calculer le gradient et la hessienne de f en tout point (x,y) de $\mathbb{R}^2$.

(2) Déterminer l'ensemble des points stationnaires de f.

(3) Pour chaque point stationnaire, dire s'il s'agit d'un minimiseur local de f, d'un minimiseur local strict de f.

(1) La fonction f est de classe $\mathscr{C}^2(\mathbb{R}^2)$.

(1) La fonction f est de classe $\mathscr{C}^2(\mathbb{R}^2)$. Pour tout $(x,y) \in \mathbb{R}^2$,

$$\nabla f(x,y) = \begin{pmatrix} 3(x+y)^2 + \operatorname{sh}(x-y) \\ 3(x+y)^2 - \operatorname{sh}(x-y) \end{pmatrix}$$

et

$$\nabla^2 f(x,y) = \begin{bmatrix} 6(x+y) + \operatorname{ch}(x-y) & 6(x+y) - \operatorname{ch}(x-y) \\ 6(x+y) - \operatorname{ch}(x-y) & 6(x+y) + \operatorname{ch}(x-y) \end{bmatrix}.$$

(2) On a:

$$\nabla f(x,y) = \mathbf{0} \iff \left\{ \begin{array}{rcl} 0 & = & 3(x+y)^2 + \mathrm{sh}(x-y) \\ 0 & = & 3(x+y)^2 - \mathrm{sh}(x-y) \end{array} \right.$$

$$\iff (x+y)^2 = 0$$

$$\iff x = y = 0.$$

(2) On a:

$$\nabla f(x,y) = \mathbf{0} \iff \left\{ \begin{array}{lcl} 0 & = & 3(x+y)^2 + \operatorname{sh}(x-y) \\ 0 & = & 3(x+y)^2 - \operatorname{sh}(x-y) \end{array} \right.$$

$$\iff (x+y)^2 = 0$$

$$\iff x = y = 0.$$

Donc l'ensemble des points stationnaires de f est $\{(0,0)\}$.

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

On vérifie que $\nabla^2 f(0,0)$ est semi-définie positive mais non définie positive.

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

On vérifie que $\nabla^2 f(0,0)$ est semi-définie positive mais non définie positive. Pour déterminer la nature du point stationnaire $(0,0)$, il faut faire une étude locale.

(3) D'après la question (1),

$$\nabla^2 f(0,0) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

On vérifie que $\nabla^2 f(0,0)$ est semi-définie positive mais non définie positive. Pour déterminer la nature du point stationnaire $(0,0)$, il faut faire une étude locale. On remarque que, pour tout $t > 0$,

$$f\left((0,0)^\top + t(-1,-1)^\top\right) = -8t^3 + 1 < 1 = f(0,0),$$

ce qui montre que $(0,0)$ n'est pas un minimum local.

Exercice

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = (x + \sin y)^2 + \sin y$.

(1) Calculer le gradient et la hessienne de f en tout point (x,y) de $\mathbb{R}^2$.

(2) Déterminer l'ensemble des points stationnaires de f.

(3) Pour chaque point stationnaire, dire s'il s'agit d'un minimiseur local de f, d'un minimiseur local strict de f.

(1) On vérifie sans peine que f est de classe $\mathscr{C}^2(\mathbb{R}^2)$ et que, pour tout $(x,y) \in \mathbb{R}^2$,

$$\nabla f(x,y) = \begin{pmatrix} 2(x+\sin y) \\ 2(x+\sin y)\cos y + \cos y \end{pmatrix}$$

et

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 2\cos y \\ 2\cos y & 2(\cos^2 y - \sin^2 y) - (2x+1)\sin y \end{bmatrix}.$$

(2) On a:

$$\nabla f(x,y) = \mathbf{0} \iff \left\{ \begin{array}{rcl} 0 & = & 2(x+\sin y) \\ 0 & = & 2(x+\sin y)\cos y + \cos y \end{array} \right.$$

$$\iff \left\{ \begin{array}{rcl} x & = & -\sin y \\ 0 & = & \cos y \end{array} \right.$$

(2) On a:

$$\nabla f(x,y) = \mathbf{0} \iff \left\{ \begin{array}{rcl} 0 & = & 2(x+\sin y) \\ 0 & = & 2(x+\sin y)\cos y + \cos y \end{array} \right.$$

$$\iff \left\{ \begin{array}{rcl} x & = & -\sin y \\ 0 & = & \cos y \end{array} \right.$$

L'ensemble des points stationnaires est donc

$$\mathscr{S} = \left\{ \left((-1)^{k+1}, \frac{\pi}{2} + k\pi \right) \,\middle|\, k \in \mathbb{Z} \right\}.$$

(3) On a:

$$\nabla^2 f\left((-1)^{k+1}, \frac{\pi}{2} + k\pi\right) = \begin{bmatrix} 2 & 0 \\ 0 & (-1)^{k+1} \end{bmatrix}.$$

(3) On a:

$$\nabla^2 f\left((-1)^{k+1}, \frac{\pi}{2}+k\pi\right) = \begin{bmatrix} 2 & 0 \\ 0 & (-1)^{k+1} \end{bmatrix}.$$

Cette matrice est définie positive pour k impair, mais ni semi-définie positive ni semi-définie négative pour k pair. Il s'ensuit que, pour k impair, $((-1)^{k+1}, \pi/2+k\pi)$ est un minimum local strict et que, pour k pair, $((-1)^{k+1}, \pi/2+k\pi)$ n'est ni un minimum local, ni un maximum local.

Exercice

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = \mathrm{ch}(y - x^2)$.

(1) Calculer le gradient et la hessienne de f en tout point (x,y) de $\mathbb{R}^2$.

(2) Déterminer l'ensemble des points stationnaires de f.

(3) Pour chaque point stationnaire, dire s'il s'agit d'un minimiseur local de f, d'un minimiseur local strict de f.

(1) On vérifie sans peine que f est de classe $\mathscr{C}^2(\mathbb{R}^2)$ et que, pour tout $(x,y) \in \mathbb{R}^2$,

$$\nabla f(x,y) = \begin{pmatrix} -2x\,\mathrm{sh}(y-x^2) \\ \mathrm{sh}(y-x^2) \end{pmatrix}$$

et

$$\nabla^2 f(x,y)$$

$$= \begin{bmatrix} 2\big(2x^2\,\mathrm{ch}(y-x^2) - \mathrm{sh}(y-x^2)\big) & -2x\,\mathrm{ch}(y-x^2) \\ -2x\,\mathrm{ch}(y-x^2) & \mathrm{ch}(y-x^2) \end{bmatrix}.$$

(2) L'ensemble des points stationnaires de f est l'ensemble

$$\{\,(x,y)\in\mathbb{R}^2 \mid \operatorname{sh}(y-x^2)=0\,\}=\{\,(x,y)\in\mathbb{R}^2 \mid y-x^2=0\,\}.$$

(2) L'ensemble des points stationnaires de f est l'ensemble

$$\left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \operatorname{sh}(y-x^2)=0 \right\} = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; y-x^2=0 \right\}.$$

(3) En tout point stationnaire (x,y) de f,

$$\nabla^2 f(x,y) = \begin{bmatrix} 4x^2 & -2x \\ -2x & 1 \end{bmatrix}.$$

(2) L'ensemble des points stationnaires de f est l'ensemble

$$\{\,(x,y)\in\mathbb{R}^2 \mid \operatorname{sh}(y-x^2)=0\,\} = \{\,(x,y)\in\mathbb{R}^2 \mid y-x^2=0\,\}\,.$$

(3) En tout point stationnaire (x,y) de f,

$$\nabla^2 f(x,y) = \begin{bmatrix} 4x^2 & -2x \\ -2x & 1 \end{bmatrix}.$$

Cette matrice est semi-définie positive, mais non définie positive.

(2) L'ensemble des points stationnaires de f est l'ensemble

$$\{ (x,y) \in \mathbb{R}^2 \mid \operatorname{sh}(y - x^2) = 0 \} = \{ (x,y) \in \mathbb{R}^2 \mid y - x^2 = 0 \}.$$

(3) En tout point stationnaire (x,y) de f,

$$\nabla^2 f(x,y) = \begin{bmatrix} 4x^2 & -2x \\ -2x & 1 \end{bmatrix}.$$

Cette matrice est semi-définie positive, mais non définie positive. Toutefois, il est clair que f est minorée par 1, et comme $f(x,y) = 1$ pour tout point stationnaire (x,y), chacun d'entre eux est un minimiseur local (non strict) de f.

Existence et conditions d'optimalité

Algorithmes

Exercice

Soient A une matrice réelle symétrique définie positive de taille $n \times n$, $\mathbf{b}$ un vecteur de $\mathbb{R}^n$ et $f\colon \mathbb{R}^n \to \mathbb{R}$ la fonction définie par

$$f(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle + \langle \mathbf{b}, \mathbf{x}\rangle.$$

On cherche à minimiser f sur $\mathbb{R}^n$. On rappelle que f admet pour unique minimiseur global $\bar{\mathbf{x}} = -A^{-1}\mathbf{b}$. L'algorithme du *gradient à pas constant* consiste à choisir un point initial $\mathbf{x}_0$ puis à construire la suite $(\mathbf{x}_k)$ *via* l'itération

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{d}_k,$$

où $\alpha > 0$ est fixé et $\mathbf{d}_k$ est la direction de plus profonde descente au point $\mathbf{x}_k$.

(1) Montrer que $\mathbf{e}_k := \mathbf{x}_k - \bar{\mathbf{x}}$ satisfait $\mathbf{e}_k = (I - \alpha A)^k \mathbf{e}_0$.

(2) En déduire que si $0 < \alpha < 2/\rho(A)$, où $\rho(A)$ est le rayon spectral de A, la suite $\mathbf{x}_k$ converge vers $\bar{\mathbf{x}}$.

(1) Le gradient de f est donné par: $\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$.

(1) Le gradient de f est donné par: $\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. Il s'ensuit que l'itération de l'algorithme du gradient à pas constant s'écrit, dans le cas présent,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha(A\mathbf{x}_k + \mathbf{b}) = (I - \alpha A)\mathbf{x}_k - \alpha\mathbf{b}.$$

(1) Le gradient de f est donné par: $\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. Il s'ensuit que l'itération de l'algorithme du gradient à pas constant s'écrit, dans le cas présent,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha(A\mathbf{x}_k + \mathbf{b}) = (I - \alpha A)\mathbf{x}_k - \alpha \mathbf{b}.$$

On a donc, pour tout $k \geq 1$,

$$\begin{aligned}
\mathbf{e}_k &= \mathbf{x}_k - \bar{\mathbf{x}} \\
&= (I - \alpha A)\mathbf{x}_{k-1} - \alpha \mathbf{b} - \bar{\mathbf{x}} + \alpha(A\bar{\mathbf{x}} + \mathbf{b}) \\
&= (I - \alpha A)(\mathbf{x}_{k-1} - \bar{\mathbf{x}}) \\
&= (I - \alpha A)\mathbf{e}_{k-1},
\end{aligned}$$

où la deuxième égalité s'appuie sur le fait que $\nabla f(\bar{\mathbf{x}}) = A\bar{\mathbf{x}} + \mathbf{b} = \mathbf{0}$ (condition nécessaire d'optimalité d'ordre 1).

(1) Le gradient de f est donné par: $\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. Il s'ensuit que l'itération de l'algorithme du gradient à pas constant s'écrit, dans le cas présent,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha(A\mathbf{x}_k + \mathbf{b}) = (I - \alpha A)\mathbf{x}_k - \alpha\mathbf{b}.$$

On a donc, pour tout $k \geq 1$,

$$\begin{aligned} \mathbf{e}_k &= \mathbf{x}_k - \bar{\mathbf{x}} \\ &= (I - \alpha A)\mathbf{x}_{k-1} - \alpha\mathbf{b} - \bar{\mathbf{x}} + \alpha(A\bar{\mathbf{x}} + \mathbf{b}) \\ &= (I - \alpha A)(\mathbf{x}_{k-1} - \bar{\mathbf{x}}) \\ &= (I - \alpha A)\mathbf{e}_{k-1}, \end{aligned}$$

où la deuxième égalité s'appuie sur le fait que $\nabla f(\bar{\mathbf{x}}) = A\bar{\mathbf{x}} + \mathbf{b} = \mathbf{0}$ (condition nécessaire d'optimalité d'ordre 1). Par une récurrence évidente, on obtient donc la relation $\mathbf{e}_k = (I - \alpha A)^k \mathbf{e}_0$.

(2) La matrice A étant symétrique définie positive, elle se diagonalise dans une base orthonormale, et ses valeurs propres sont toutes strictement positives.

(2) La matrice A étant symétrique définie positive, elle se diagonalise dans une base orthonormale, et ses valeurs propres sont toutes strictement positives. Autrement dit, il existe une base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$ et des réels strictements positifs $\lambda_1, \ldots, \lambda_n$ tels que

$$\forall j \in \{1, \ldots, n\}, \quad A\mathbf{v}_j = \lambda_j \mathbf{v}_j.$$

(2) La matrice A étant symétrique définie positive, elle se diagonalise dans une base orthonormale, et ses valeurs propres sont toutes strictement positives. Autrement dit, il existe une base orthonormale $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$ et des réels strictement positifs $\lambda_1, \ldots, \lambda_n$ tels que

$$\forall j \in \{1, \ldots, n\}, \quad A\mathbf{v}_j = \lambda_j \mathbf{v}_j.$$

Si l'on note $\langle \cdot, \cdot \rangle$ le produit scalaire euclidien de $\mathbb{R}^n$, on a alors, pour tout $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{x}, \mathbf{v}_j \rangle \mathbf{v}_j, \quad \text{et donc} \quad (I - \alpha A)\mathbf{x} = \sum_{j=1}^{n} (1 - \alpha\lambda_j)\xi_j \mathbf{v}_j,$$

en posant $\xi_j := \langle \mathbf{x}, \mathbf{v}_j \rangle$.

On en déduit que

$$\|(I-\alpha A)\mathbf{x}\|^2 = \sum_{j=1}^{n}(1-\alpha\lambda_j)^2\xi_j^2$$

On en déduit que

$$\begin{aligned}\|(I-\alpha A)\mathbf{x}\|^2 &= \sum_{j=1}^{n}(1-\alpha\lambda_j)^2\xi_j^2\\ &\leq \max_{1\leq j\leq n}(1-\alpha\lambda_j)^2\cdot\sum_{j=1}^{n}\xi_j^2\end{aligned}$$

On en déduit que

$$
\begin{aligned}
\|(I-\alpha A)\mathbf{x}\|^2 &= \sum_{j=1}^{n}(1-\alpha\lambda_j)^2\xi_j^2 \\
&\leq \max_{1\leq j\leq n}(1-\alpha\lambda_j)^2\cdot\sum_{j=1}^{n}\xi_j^2 \\
&= \max_{1\leq j\leq n}(1-\alpha\lambda_j)^2\cdot\|\mathbf{x}\|^2.
\end{aligned}
$$

On en déduit que

$$\begin{aligned}
\|(I-\alpha A)\mathbf{x}\|^2 &= \sum_{j=1}^{n}(1-\alpha\lambda_j)^2\xi_j^2 \\
&\leq \max_{1\leq j\leq n}(1-\alpha\lambda_j)^2\cdot\sum_{j=1}^{n}\xi_j^2 \\
&= \max_{1\leq j\leq n}(1-\alpha\lambda_j)^2\cdot\|\mathbf{x}\|^2.
\end{aligned}$$

On obtient donc que, pour tout $\mathbf{x}\in\mathbb{R}^n$,

$$\|(I-\alpha A)\mathbf{x}\| \leq \max_j|1-\alpha\lambda_j|\cdot\|\mathbf{x}\|.$$

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha \lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha \lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha \lambda_j| < 1$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha \lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha \lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha \lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$. De plus,

$$\max_j |1 - \alpha\lambda_j| = \max\left\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\right\}.$$

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$. De plus,

$$\max_j |1 - \alpha\lambda_j| = \max\left\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\right\}.$$

Si $0 < \alpha < 2/\rho$, alors $0 < \alpha\lambda_{\max} < 2$ et donc $0 < \alpha\lambda_{\min} < 2$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$. De plus,

$$\max_j |1 - \alpha\lambda_j| = \max\left\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\right\}.$$

Si $0 < \alpha < 2/\rho$, alors $0 < \alpha\lambda_{\max} < 2$ et donc $0 < \alpha\lambda_{\min} < 2$. De ces inégalités, on tire aisément que $|1 - \alpha\lambda_{\max}| < 1$ et $|1 - \alpha\lambda_{\min}| < 1$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$. De plus,

$$\max_j |1 - \alpha\lambda_j| = \max\left\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\right\}.$$

Si $0 < \alpha < 2/\rho$, alors $0 < \alpha\lambda_{\max} < 2$ et donc $0 < \alpha\lambda_{\min} < 2$. De ces inégalités, on tire aisément que $|1 - \alpha\lambda_{\max}| < 1$ et $|1 - \alpha\lambda_{\min}| < 1$. Il s'ensuit que $\max_j |1 - \alpha\lambda_j| < 1$.

Par une récurrence évidente, on en déduit que, pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $k \in \mathbb{N}$,

$$\|(I - \alpha A)^k \mathbf{x}\| \leq \left(\max_j |1 - \alpha\lambda_j|\right)^k \cdot \|\mathbf{x}\|.$$

En particulier, pour $\mathbf{x} = \mathbf{e}_0$, on obtient

$$\|\mathbf{e}_k\| \leq (\max_j |1 - \alpha\lambda_j|)^k \cdot \|\mathbf{e}_0\|. \tag{2}$$

Montrons que $\max_j |1 - \alpha\lambda_j| < 1$. Posons $\lambda_{\max} := \max_j \lambda_j$ et $\lambda_{\min} := \min_j \lambda_j$. Il est clair que, dans le cas présent, $\rho(A) = \lambda_{\max}$. De plus,

$$\max_j |1 - \alpha\lambda_j| = \max \left\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\right\}.$$

Si $0 < \alpha < 2/\rho$, alors $0 < \alpha\lambda_{\max} < 2$ et donc $0 < \alpha\lambda_{\min} < 2$. De ces inégalités, on tire aisément que $|1 - \alpha\lambda_{\max}| < 1$ et $|1 - \alpha\lambda_{\min}| < 1$. Il s'ensuit que $\max_j |1 - \alpha\lambda_j| < 1$. L'inégalité (2) montre alors que $\mathbf{x}_k \to \bar{\mathbf{x}}$ lorsque $k \to \infty$.

Exercice

Soit $f : \mathbb{R}^p \to \mathbb{R}$ une fonction continûment différentiable de gradient L-lipschitzien ($L > 0$). On rappelle que cela signifie que, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| \leq L\|\mathbf{y} - \mathbf{x}\|.$$

(1) Justifier que pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned} & f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\ = & \int_0^1 \big\langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \big\rangle \, \mathrm{d}t. \end{aligned}$$

(2) En déduire que, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

(1) Posons $\varphi(t) = f\big((1-t)\mathbf{x} + t\mathbf{y}\big) = f\big(\mathbf{x} + t(\mathbf{y}-\mathbf{x})\big)$, de sorte que $f(\mathbf{y}) = \varphi(1)$ et $f(\mathbf{x}) = \varphi(0)$.

(1) Posons $\varphi(t) = f\big((1-t)\mathbf{x} + t\mathbf{y}\big) = f\big(\mathbf{x} + t(\mathbf{y}-\mathbf{x})\big)$, de sorte que $f(\mathbf{y}) = \varphi(1)$ et $f(\mathbf{x}) = \varphi(0)$. La fonction f étant de classe $\mathscr{C}^1$, il en est de même de φ et, en posant $\mathbf{u}(t) = \mathbf{x} + t(\mathbf{y}-\mathbf{x})$,

$$\varphi'(t) = \sum_{k=1}^{p} \frac{\partial f}{\partial u_k}(\mathbf{u}(t)) \cdot \frac{\partial u_k}{\partial t}(t) = \big\langle \nabla f\big(\mathbf{u}(t)\big), \mathbf{y}-\mathbf{x} \big\rangle .$$

D’après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$f(\mathbf{y}) - f(\mathbf{x}) \; = \; \varphi(1) - \varphi(0)$$

D'après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= \varphi(1) - \varphi(0) \\ &= \int_0^1 \varphi'(t)\mathrm{d}t \end{aligned}$$

D'après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= \varphi(1) - \varphi(0) \\ &= \int_0^1 \varphi'(t)\mathrm{d}t \\ &= \int_0^1 \left\langle \nabla f\left(\mathbf{u}(t)\right), \mathbf{y} - \mathbf{x} \right\rangle \mathrm{d}t \end{aligned}$$

D'après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned}
f(\mathbf{y}) - f(\mathbf{x}) &= \varphi(1) - \varphi(0) \\
&= \int_0^1 \varphi'(t)\mathrm{d}t \\
&= \int_0^1 \left\langle \nabla f\left(\mathbf{u}(t)\right), \mathbf{y} - \mathbf{x} \right\rangle \mathrm{d}t \\
&= \int_0^1 \langle \nabla f((1-t)\mathbf{x} + t\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t
\end{aligned}$$

D'après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= \varphi(1) - \varphi(0) \\ &= \int_0^1 \varphi'(t) \mathrm{d}t \\ &= \int_0^1 \left\langle \nabla f\big(\mathbf{u}(t)\big), \mathbf{y} - \mathbf{x} \right\rangle \mathrm{d}t \\ &= \int_0^1 \langle \nabla f((1-t)\mathbf{x} + t\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\ &= \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}) + \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \end{aligned}$$

D'après le théorème fondamental du calcul intégral, nous avons alors, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned}
f(\mathbf{y}) - f(\mathbf{x}) &= \varphi(1) - \varphi(0) \\
&= \int_0^1 \varphi'(t)\mathrm{d}t \\
&= \int_0^1 \left\langle \nabla f\big(\mathbf{u}(t)\big)\,,\, \mathbf{y} - \mathbf{x} \right\rangle \mathrm{d}t \\
&= \int_0^1 \langle \nabla f((1-t)\mathbf{x} + t\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\
&= \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}) + \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\
&= \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t(\mathbf{y})\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t.
\end{aligned}$$

(2) On en déduit que

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$

(2) On en déduit que

$$
\begin{aligned}
& f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\
= \quad & \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t
\end{aligned}
$$

(2) On en déduit que

$$
\begin{aligned}
& f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\
& \quad = \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\
& \quad \leq \int_0^1 \|\nabla f((1-t)\mathbf{x} + t\mathbf{y}) - \nabla f(\mathbf{x})\| \times \|\mathbf{y} - \mathbf{x}\| \mathrm{d}t
\end{aligned}
$$

(2) On en déduit que

$$
\begin{aligned}
& f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\
= & \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\
\leq & \int_0^1 \| \nabla f((1-t)\mathbf{x} + t\mathbf{y}) - \nabla f(\mathbf{x}) \| \times \| \mathbf{y} - \mathbf{x} \| \mathrm{d}t \\
\leq & \int_0^1 tL \times \| \mathbf{y} - \mathbf{x} \|^2 \mathrm{d}t
\end{aligned}
$$

(2) On en déduit que

$$\begin{aligned}
& f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\
&\quad = \int_0^1 \langle \nabla f\big((1-t)\mathbf{x} + t\mathbf{y}\big) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \mathrm{d}t \\
&\quad \leq \int_0^1 \|\nabla f((1-t)\mathbf{x} + t\mathbf{y}) - \nabla f(\mathbf{x})\| \times \|\mathbf{y} - \mathbf{x}\| \mathrm{d}t \\
&\quad \leq \int_0^1 tL \times \|\mathbf{y} - \mathbf{x}\|^2 \mathrm{d}t \\
&\quad = \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2,
\end{aligned}$$

et le résultat s'ensuit.

Exercice

Soit $f : \mathbb{R}^p \to \mathbb{R}$ une fonction continûment différentiable de gradient L-lipschitzien, et telle que $\inf_{\mathbf{x}\in\mathbb{R}^p} f(\mathbf{x}) > -\infty$. On considère l'algorithme suivant, pour $\mathbf{x}_0 \in \mathbb{R}^p$, $s > 0$ et $k \in \mathbb{N}$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - s\nabla f(\mathbf{x}_k). \tag{3}$$

(1) En utilisant l'exercice précédent, montrer que pour tout $k \in \mathbb{N}$

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - s\left(1 - \frac{sL}{2}\right)\|\nabla f(\mathbf{x}_k)\|^2.$$

(2) En déduire que si $s \in (0, 2/L)$, alors $(f(\mathbf{x}_k))_{k\in\mathbb{N}}$ converge et $\|\nabla f(\mathbf{x}_k)\|$ tend vers 0 lorsque $k \to \infty$.

(1) D'après l'exercice précédent, pour tout $k \in \mathbb{N}$,

$$\begin{aligned} & f(\mathbf{x}_{k+1}) \\ & \quad \leq \quad f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \end{aligned}$$

(1) D'après l'exercice précédent, pour tout $k \in \mathbb{N}$,

$$\begin{aligned} & f(\mathbf{x}_{k+1}) \\ & \quad \leq \quad f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \\ & \quad = \quad f(\mathbf{x}_k) - \|\nabla f(\mathbf{x}_k)\|^2 \left(s - \frac{s^2 L}{2} \right) \end{aligned}$$

(1) D'après l'exercice précédent, pour tout $k \in \mathbb{N}$,

$$
\begin{aligned}
& f(\mathbf{x}_{k+1}) \\
& \quad \leq \quad f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \\
& \quad = \quad f(\mathbf{x}_k) - \|\nabla f(\mathbf{x}_k)\|^2 \left(s - \frac{s^2 L}{2} \right) \\
& \quad = \quad f(\mathbf{x}_k) - s \left(1 - \frac{sL}{2} \right) \|\nabla f(\mathbf{x}_k)\|^2.
\end{aligned}
$$

(2) Si $s \in (0, 2/L)$, alors

$$\alpha := s\left(1 - \frac{sL}{2}\right) > 0$$

et la suite $\big(f(\mathbf{x}_k)\big)$ est décroissante.

(2) Si $s \in (0, 2/L)$, alors

$$\alpha := s\left(1 - \frac{sL}{2}\right) > 0$$

et la suite $\big(f(\mathbf{x}_k)\big)$ est décroissante. Puisqu'elle est bornée inférieurement, elle est convergente.

(2) Si $s \in (0, 2/L)$, alors

$$\alpha := s\left(1 - \frac{sL}{2}\right) > 0$$

et la suite $\big(f(\mathbf{x}_k)\big)$ est décroissante. Puisqu'elle est bornée inférieurement, elle est convergente. En sommant les inégalités

$$\|\nabla f(\mathbf{x}_k)\|^2 \leq \frac{1}{\alpha}\big(f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})\big)$$

pour k allant de 0 à n, on obtient par télescopage:

$$\sum_{k=0}^{n} \|\nabla f(\mathbf{x}_k)\|^2 \leq \frac{1}{\alpha}\big(f(\mathbf{x}_0) - f(\mathbf{x}_{n+1})\big) \leq \frac{1}{\alpha}\left(f(\mathbf{x}_0) - \inf_{\mathbf{x}\in\mathbb{R}^p} f(\mathbf{x})\right)$$

ce qui montre que la série positive est sommable, et donc que son terme général converge vers 0.

Exercice

Soit $f : \mathbb{R}^p \to \mathbb{R}$ une fonction continûment différentiable de gradient L-lipschitzien, et telle que $\bar{f} := \inf_{\mathbf{x} \in \mathbb{R}^p} f(\mathbf{x}) > -\infty$. On suppose de plus que f est convexe, et qu'elle atteint son minimum en un point $\bar{\mathbf{x}}$. Soit un point initial $\mathbf{x}_0 \in \mathbb{R}^p$ et l'itération

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla f(\mathbf{x}_k).$$

(1) Vérifier que, pour tout $k \in \mathbb{N}$ et tout $\mathbf{y} \in \mathbb{R}^p$,

$$\begin{aligned} & f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{y} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_k\|^2 \\ &= f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_{k+1}\|^2. \end{aligned}$$

(2) Montrer que pour tout $k \in \mathbb{N}$ et tout $\bar{\mathbf{x}} \in \operatorname{argmin} f$,

$$f(\bar{\mathbf{x}}) + \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_k\|^2 \geq f(\mathbf{x}_{k+1}) + \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2.$$

(3) Déduire de l'inégalité précédente que, pour tout $K \in \mathbb{N}^*$ et tout $\bar{\mathbf{x}} \in \operatorname{argmin} f$,

$$f(\mathbf{x}_K) - \bar{f} \leq \frac{L}{2K}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

Que peut-on en déduire quant à la convergence de la suite $\big(f(\mathbf{x}_k)\big)$?

(4) Déduire de ce qui précède que la suite $(\mathbf{x}_K)_K$ est bornée, puis qu'elle est convergente.

(1) On remarque que $\nabla f(\mathbf{x}_k) = L(\mathbf{x}_k - \mathbf{x}_{k+1})$. Alors

$$f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_{k+1}\|^2$$

(1) On remarque que $\nabla f(\mathbf{x}_k) = L(\mathbf{x}_k - \mathbf{x}_{k+1})$. Alors

$$
\begin{aligned}
& f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_{k+1}\|^2 \\
= \quad & f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
& + \frac{L}{2}\left(\|\mathbf{y} - \mathbf{x}_k\|^2 + 2\langle \mathbf{y} - \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle + \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2\right)
\end{aligned}
$$

(1) On remarque que $\nabla f(\mathbf{x}_k) = L(\mathbf{x}_k - \mathbf{x}_{k+1})$. Alors

$$
\begin{aligned}
& f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_{k+1}\|^2 \\
= \quad & f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
& + \frac{L}{2}\left(\|\mathbf{y} - \mathbf{x}_k\|^2 + 2\langle \mathbf{y} - \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle + \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2\right) \\
= \quad & f(\mathbf{x}_k) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_k\|^2 + L\langle \mathbf{y} - \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle
\end{aligned}
$$

(1) On remarque que $\nabla f(\mathbf{x}_k) = L(\mathbf{x}_k - \mathbf{x}_{k+1})$. Alors

$$
\begin{aligned}
& f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_{k+1}\|^2 \\
= \quad & f(\mathbf{x}_k) + \left\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \right\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
& + \frac{L}{2}\left(\|\mathbf{y} - \mathbf{x}_k\|^2 + 2\langle \mathbf{y} - \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle + \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2\right) \\
= \quad & f(\mathbf{x}_k) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_k\|^2 + L\langle \mathbf{y} - \mathbf{x}_k, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle \\
= \quad & f(\mathbf{x}_k) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}_k\|^2 + \langle \mathbf{y} - \mathbf{x}_k, \nabla f(\mathbf{x}_k)\rangle.
\end{aligned}
$$

(2) On a montré dans un exercice précédent, que pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2.$$

(2) On a montré dans un exercice précédent, que pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2.$$

En prenant $\mathbf{y} = \mathbf{x}_{k+1}$ et $\mathbf{x} = \mathbf{x}_k$, on obtient:

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) - \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle \leq \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

(2) On a montré dans un exercice précédent, que pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$,

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2.$$

En prenant $\mathbf{y} = \mathbf{x}_{k+1}$ et $\mathbf{x} = \mathbf{x}_k$, on obtient:

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) - \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle \leq \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

On en déduit que

$$\begin{aligned}
& f(\mathbf{x}_{k+1}) + \frac{L}{2} \|\mathbf{x}_{k+1} - \bar{\mathbf{x}}\|^2 \\
& \quad \leq \quad f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \frac{L}{2} \|\mathbf{x}_{k+1} - \bar{\mathbf{x}}\|^2 \\
& \quad = \quad f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \bar{\mathbf{x}} - \mathbf{x}_k \rangle + \frac{L}{2} \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2 \\
& \quad \leq \quad f(\bar{\mathbf{x}}) + \frac{L}{2} \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2,
\end{aligned}$$

où l'égalité provient de la question (1) avec $\mathbf{y} = \bar{\mathbf{x}}$ et l'inégalité qui suit de la convexité de f.

(3) D'après (2), on a:

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1}) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2\right\}.$$

(3) D'après (2), on a:

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1}) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2\right\}.$$

En sommant ces inégalités pour $k \in \{0, \ldots, K-1\}$, on obtient par télescopage

$$\sum_{k=0}^{K-1} \big(f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1})\big) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_0\|^2\right\}.$$

(3) D'après (2), on a:

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1}) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2\right\}.$$

En sommant ces inégalités pour $k \in \{0, \ldots, K-1\}$, on obtient par télescopage

$$\sum_{k=0}^{K-1}\big(f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1})\big) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_0\|^2\right\}.$$

Il s'ensuit que

$$\sum_{k=0}^{K-1}\big(f(\mathbf{x}_{k+1}) - f(\bar{\mathbf{x}})\big) \leq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2\right\} \leq \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

(3) D'après (2), on a:

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1}) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2\right\}.$$

En sommant ces inégalités pour $k \in \{0, \ldots, K-1\}$, on obtient par télescopage

$$\sum_{k=0}^{K-1}\left(f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1})\right) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_0\|^2\right\}.$$

Il s'ensuit que

$$\sum_{k=0}^{K-1}\left(f(\mathbf{x}_{k+1}) - f(\bar{\mathbf{x}})\right) \leq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2\right\} \leq \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

Puisque la suite $\left(f(\mathbf{x}_k)\right)$ est décroissante, on a:

$$K\left(f(\mathbf{x}_K) - f(\bar{\mathbf{x}})\right) \leq \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

(3) D'après (2), on a:

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1}) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_{k+1}\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_k\|^2\right\}.$$

En sommant ces inégalités pour $k \in \{0, \ldots, K-1\}$, on obtient par télescopage

$$\sum_{k=0}^{K-1}\left(f(\bar{\mathbf{x}}) - f(\mathbf{x}_{k+1})\right) \geq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_0\|^2\right\}.$$

Il s'ensuit que

$$\sum_{k=0}^{K-1}\left(f(\mathbf{x}_{k+1}) - f(\bar{\mathbf{x}})\right) \leq \frac{L}{2}\left\{\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2 - \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2\right\} \leq \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

Puisque la suite $\left(f(\mathbf{x}_k)\right)$ est décroissante, on a:

$$K\left(f(\mathbf{x}_K) - f(\bar{\mathbf{x}})\right) \leq \frac{L}{2}\|\bar{\mathbf{x}} - \mathbf{x}_0\|^2.$$

Ainsi, $f(\mathbf{x}_K) \to f(\bar{\mathbf{x}}) = \bar{f}$ lorsque $K \to \infty$ avec un taux $O(1/K)$.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}}-\mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K)-f(\bar{\mathbf{x}})\big)+\|\bar{\mathbf{x}}-\mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}}-\mathbf{x}_K\|^2.$$

(4) D'après la question (2),

$$\|\bar{\mathbf{x}}-\mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K)-f(\bar{\mathbf{x}})\big)+\|\bar{\mathbf{x}}-\mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}}-\mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}}-\mathbf{x}_K\|\big)_K$ est décroissante.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}}-\mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K)-f(\bar{\mathbf{x}})\big)+\|\bar{\mathbf{x}}-\mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}}-\mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}}-\mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}}-\mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K)-f(\bar{\mathbf{x}})\big)+\|\bar{\mathbf{x}}-\mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}}-\mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}}-\mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée. Il s'ensuit que la suite $(\mathbf{x}_K)_K$ est elle-même bornée.

La suite $(\mathbf{x}_K)_{K\geq 1}$ admet donc un point d'accumulation $\check{\mathbf{x}}$: on a une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$ et par continuité de f on a aussi que $f(\mathbf{x}_{K_j}) \to f(\check{\mathbf{x}})$.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}}-\mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K)-f(\bar{\mathbf{x}})\big)+\|\bar{\mathbf{x}}-\mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}}-\mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}}-\mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée. Il s'ensuit que la suite $(\mathbf{x}_K)_K$ est elle-même bornée.

La suite $(\mathbf{x}_K)_{K\geq 1}$ admet donc un point d'accumulation $\check{\mathbf{x}}$: on a une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$ et par continuité de f on a aussi que $f(\mathbf{x}_{K_j}) \to f(\check{\mathbf{x}})$. On a donc: $f(\check{\mathbf{x}}) = f(\bar{\mathbf{x}}) = \bar{f}$, de sorte que $\check{\mathbf{x}} \in \operatorname{argmin} f$.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}} - \mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K) - f(\bar{\mathbf{x}})\big) + \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}} - \mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée. Il s'ensuit que la suite $(\mathbf{x}_K)_K$ est elle-même bornée.

La suite $(\mathbf{x}_K)_{K\geq 1}$ admet donc un point d'accumulation $\check{\mathbf{x}}$: on a une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$ et par continuité de f on a aussi que $f(\mathbf{x}_{K_j}) \to f(\check{\mathbf{x}})$. On a donc: $f(\check{\mathbf{x}}) = f(\bar{\mathbf{x}}) = \bar{f}$, de sorte que $\check{\mathbf{x}} \in \operatorname{argmin} f$. Il s'ensuit, par la question (2) et la première partie de cette question, que la suite $\big(\|\check{\mathbf{x}} - \mathbf{x}_K\|\big)$ est décroissante.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}} - \mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K) - f(\bar{\mathbf{x}})\big) + \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}} - \mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée. Il s'ensuit que la suite $(\mathbf{x}_K)_K$ est elle-même bornée.

La suite $(\mathbf{x}_K)_{K\geq 1}$ admet donc un point d'accumulation $\check{\mathbf{x}}$: on a une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$ et par continuité de f on a aussi que $f(\mathbf{x}_{K_j}) \to f(\check{\mathbf{x}})$. On a donc: $f(\check{\mathbf{x}}) = f(\bar{\mathbf{x}}) = \bar{f}$, de sorte que $\check{\mathbf{x}} \in \operatorname{argmin} f$. Il s'ensuit, par la question (2) et la première partie de cette question, que la suite $\big(\|\check{\mathbf{x}} - \mathbf{x}_K\|\big)$ est décroissante. Puisque $\check{\mathbf{x}}$ est un point d'accumulation de $(\mathbf{x}_K)$, il existe une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$.

(4) D'après la question (2),

$$\|\bar{\mathbf{x}} - \mathbf{x}_{K-1}\|^2 \geq \frac{2}{L}\big(f(\mathbf{x}_K) - f(\bar{\mathbf{x}})\big) + \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2 \geq \|\bar{\mathbf{x}} - \mathbf{x}_K\|^2.$$

Ainsi, la suite $\big(\|\bar{\mathbf{x}} - \mathbf{x}_K\|\big)_K$ est décroissante. Puisqu'elle est minorée par zéro, elle est convergente, donc bornée. Il s'ensuit que la suite $(\mathbf{x}_K)_K$ est elle-même bornée.

La suite $(\mathbf{x}_K)_{K\geq 1}$ admet donc un point d'accumulation $\check{\mathbf{x}}$: on a une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$ et par continuité de f on a aussi que $f(\mathbf{x}_{K_j}) \to f(\check{\mathbf{x}})$. On a donc: $f(\check{\mathbf{x}}) = f(\bar{\mathbf{x}}) = \bar{f}$, de sorte que $\check{\mathbf{x}} \in \operatorname{argmin} f$. Il s'ensuit, par la question (2) et la première partie de cette question, que la suite $\big(\|\check{\mathbf{x}} - \mathbf{x}_K\|\big)$ est décroissante. Puisque $\check{\mathbf{x}}$ est un point d'accumulation de $(\mathbf{x}_K)$, il existe une sous-suite $(\mathbf{x}_{K_j})_j$ qui converge vers $\check{\mathbf{x}}$. Ainsi, $\|\check{\mathbf{x}} - \mathbf{x}_{K_j}\| \to 0$, et puisque $\big(\|\check{\mathbf{x}} - \mathbf{x}_K\|\big)$ est décroissante, on en déduit que $\|\check{\mathbf{x}} - \mathbf{x}_K\| \to 0$, c'est-à-dire $\mathbf{x}_K \to \check{\mathbf{x}}$, lorsque $K \to \infty$.