DM n°2 facultatif

X, ÉNS

L'objet de ce problème est la minimisation sur un sous-domaine $\mathrm{K} \subset \mathbb{R}^n$ d'une fonction f définie sur $\mathbb{R}^n$. Nous proposons plusieurs approches pour trouver des conditions nécessaires d'optimalité, et obtenir des approximations des minimiseur de f dans des cas particuliers. Les dépendances entre les parties du problème sont données par le schéma suivant :

$$\boxed{\begin{array}{ccccc} & & \mathrm{II} & & \\ & \nearrow & & & \\ \mathrm{I} & & & & \\ & \searrow & & & \\ & & \mathrm{III} & \longrightarrow & \mathrm{IV} \end{array}}$$

Notations et Définitions

Pour tout entier $k \in \mathbb{N}^*$, on désignera le produit scalaire usuel sur $\mathbb{R}^k$ par $\langle \cdot, \cdot \rangle$, et la norme euclidienne sur $\mathbb{R}^k$ par $\|\cdot\|$.

Dans tout le sujet, on se place sur $\mathbb{R}^n$, où $n \in \mathbb{N}^*$.

— *On dit qu'une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est **convexe** si pour tous $x, y \in \mathbb{R}^n$ et tout $\lambda \in [0, 1]$,*

$$f((1-\lambda)x + \lambda y) \leqslant (1-\lambda)f(x) + \lambda f(y)$$

— *On dit qu'une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est **coercive** si $\lim\limits_{\|x\| \to +\infty} f(x) = +\infty$.*

I - Préliminaires

Fonctions convexes

I.1. On considère une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

(a) Pour tous $x, y \in \mathbb{R}^n$, soit $\varphi_{x,y} : \mathbb{R} \longrightarrow \mathbb{R}$ la fonction définie par

$$\forall t \in \mathbb{R},\ \varphi_{x,y}(t) = f(x + t(y-x))$$

Montrer que f est convexe si et seulement si pour tous $x, y \in \mathbb{R}^n$, $\varphi_{x,y}$ est convexe.

(b) On suppose que f est différentiable sur $\mathbb{R}^n$. Montrer que pour tous $x, y \in \mathbb{R}^n$, la fonction $\varphi_{x,y}$ est dérivable, et montrer que

$$\forall t \in \mathbb{R},\ \varphi'_{x,y}(t) = \langle \nabla f(x + t(y-x)), y - x \rangle$$

(c) En déduire que si f est différentiable sur $\mathbb{R}^n$, alors f est convexe si et seulement si pour tous $x, y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle$$

(d) Montrer que si $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est différentiable sur $\mathbb{R}^n$, alors f est convexe si et seulement si pour tous $x, y \in \mathbb{R}^n$,

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant 0$$

I.2. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction convexe et différentiable sur $\mathbb{R}^n$, et $x^* \in \mathbb{R}^n$. Montrer que si $\nabla f(x^*) = 0$ alors f admet un minimum global en x^*.

Définition. *Soit $\alpha \in \mathbb{R}_+^*$. On dit que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable sur $\mathbb{R}^n$ est **α-convexe** si pour tous $x, y \in \mathbb{R}^n$,*

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2$$

I.3. On considère un réel $\alpha > 0$ et une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable sur $\mathbb{R}^n$.

(a) On considère la fonction $g_\alpha : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par

$$\forall x \in \mathbb{R}^n,\ g_\alpha(x) = f(x) - \frac{\alpha}{2}\|x\|^2$$

Calculer $\nabla g_\alpha(x)$ pour tout $x \in \mathbb{R}^n$, et montrer que f est α-convexe si et seulement si g_α est convexe.

(b) En déduire que f est α-convexe si et seulement si pour tous $x, y \in \mathbb{R}^n$,

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant \alpha\|y - x\|^2$$

Fonctions coercives

I.4. Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et coercive. Montrer que si K est un fermé non vide de $\mathbb{R}^n$, alors il existe $x^* \in \mathrm{K}$ tel que $f(x^*) = \inf\limits_{x \in \mathrm{K}} f(x)$.

I.5. Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$ et α-convexe où $\alpha \in \mathbb{R}_+^*$. Montrer que si K est un convexe fermé non vide de $\mathbb{R}^n$, alors f admet un unique minimum sur K.

Projection sur un convexe fermé

I.6. Soient C un convexe fermé non vide de $\mathbb{R}^n$ et $x \in \mathbb{R}^n$.

(a) Montrer qu'il existe un unique point $\mathrm{P}_c(x) \in \mathrm{C}$ tel que $\|\mathrm{P}_\mathrm{C}(x) - x\| = \inf\limits_{y \in \mathrm{C}} \|y - x\|$.

(b) Soit $\overline{x} \in \mathrm{C}$. Montrer que $\overline{x} = \mathrm{P}_\mathrm{C}(x)$ si et seulement si

$$\forall y \in \mathrm{C},\ \langle x - \overline{x}, y - \overline{x} \rangle \leqslant 0$$

Indication : on pourra considérer la fonction $\psi_y : t \in \mathbb{R} \mapsto \|x - (\overline{x} + t(y - \overline{x}))\|^2$ où $y \in \mathrm{C}$.

(c) En déduire que si $x, y \in \mathbb{R}^n$, $\|\mathrm{P}_\mathrm{C}(y) - \mathrm{P}_\mathrm{C}(x)\| \leqslant \|y - x\|$.

Une première condition nécessaire d'optimalité

Soit $\mathrm{K} \subset \mathbb{R}^n$. On dit qu'un vecteur $h \in \mathbb{R}^n$ est K-admissible au point $x \in \mathrm{K}$ s'il existe

- *une suite $(t_k)_{k \in \mathbb{N}}$ de réels strictement positifs vérifiant $\lim\limits_{k \longrightarrow +\infty} t_k = 0$,*
- *une suite $(h_k)_{k \in \mathbb{N}}$ de vecteurs de $\mathbb{R}^n$ vérifiant $\lim\limits_{k \longrightarrow +\infty} h_k = h$,*

telles que pour tout $k \in \mathbb{N}$,

$$x + t_k h_k \in \mathrm{K}$$

On appelle cône K-admissible au point $x \in \mathrm{K}$ l'ensemble

$$\mathcal{A}_\mathrm{K}(x) := \{h \in \mathbb{R}^n,\ h \ \textit{ est un vecteur } \mathrm{K} \textit{ admissible au point } x\}$$

I.7. Décrire $\mathcal{A}_\mathrm{K}(x)$ dans le cas où x est dans l'intérieur de K.

I.8. Montrer que si $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est différentiable en $x^* \in \mathrm{K}$ et admet un minimum local sur K en x^*, alors

$$\forall h \in \mathcal{A}_\mathrm{K}(x^*),\ \langle \nabla f(x^*), h \rangle \geqslant 0$$

Qu'exprime ce résultat dans le cas particulier où x^* est dans l'intérieur de K ?

II - Pénalisation

Dans le but d'approcher un minimum d'une fonction f sur $\mathrm{K} \subset \mathbb{R}^n$, on cherche à se ramener à la minimisation d'une fonction sur $\mathbb{R}^n$ tout entier. Pour ce faire, on propose d'ajouter à f un terme de « pénalisation », qui prend de grandes valeurs en dehors de K, et de minimiser la nouvelle fonction pénalisée sur $\mathbb{R}^n$ tout entier. Cette partie a pour but de justifier cette approche dans un cas particulier.

Dans toute cette partie, on considère une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable sur $\mathbb{R}^n$ et α-convexe, où $\alpha \in \mathbb{R}_+^*$. On pose

$$\mathrm{K} = \{x \in \mathbb{R}^n,\ g_1(x) \leqslant 0, \ldots, g_p(x) \leqslant 0\}$$

où $p \in \mathbb{N}^*$ et $g_1, \ldots, g_p$ sont des fonctions convexes de $\mathbb{R}^n$ dans $\mathbb{R}$, différentiables sur $\mathbb{R}^n$. On suppose de plus que l'ensemble K est non vide.

II.1. Montrer qu'il existe un unique élément $x^* \in \mathrm{K}$ tel que $f(x^*) = \inf_{x \in \mathrm{K}} f(x)$.

Pour tout $k \in \mathbb{N}$, on introduit la fonction $f_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par

$$\forall x \in \mathbb{R}^n,\ f_k(x) = f(x) + k\Psi(x)$$

où $\Psi : \mathbb{R}^n \longrightarrow \mathbb{R}$ est la fonction définie par

$$\forall x \in \mathbb{R}^n,\ \Psi(x) = \sum_{i=1}^{p} \max(0, g_i(x))^2$$

II.2. Pour tout $x \in \mathbb{R}^n$, calculer $\lim_{k \to +\infty} f_k(x)$.

II.3. Montrer que pour tout $k \in \mathbb{N}$, il existe un unique $x_k \in \mathbb{R}^n$ tel que $f_k(x_k) = \inf_{x \in \mathbb{R}^n} f_k(x)$.

Indication : on pourra commencer par montrer que si $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ est une fonction convexe, et $h : \mathbb{R} \to \mathbb{R}$ est une fonction convexe croissante, alors $h \circ g$ est convexe.

II.4. Montrer que pour tout $k \in \mathbb{N}$, $f(x_k) \leqslant f(x^*)$.

II.5. On considère une sous-suite $(x_{\varphi(k)})_{k \in \mathbb{N}}$ de $(x_k)_{k \in \mathbb{N}}$ qui converge vers $\overline{x} \in \mathbb{R}^n$.

(a) Montrer que $\overline{x} \in \mathrm{K}$.

(b) En déduire que $\overline{x} = x^*$.

II.6. En déduire que la suite $(x_k)_{k \in \mathbb{N}}$ converge vers x^*.

II.7. Montrer que la suite $(f_k(x_k))_{k \in \mathbb{N}}$ converge vers $f(x^*)$.

III - Théorème de Karush-Kuhn-Tucker

Le but de cette partie est d'établir une condition nécessaire d'optimalité dans le cas où le domaine K est décrit par des contraintes de type inégalité.

Lemme de Farkas

Soient $m \in \mathbb{N}^*$ et $(u_1, \ldots, u_m)$ une famille de vecteurs de $\mathbb{R}^n$. On note

$$\mathrm{C} = \left\{ \sum_{i=1}^{m} \mu_i u_i,\ \forall i \in [\![1, m]\!]\ \mu_i \geqslant 0 \right\}$$

On cherche à démonter le résultat suivant.

Lemme 1. *Si $v \in \mathbb{R}^n$, alors une et une seule des deux assertions suivantes est vérifiée :*
(i) $v \in \mathrm{C}$,
(ii) *il existe $w \in \mathbb{R}^n$ tel que $\langle v, w\rangle < 0$ et $\forall i \in [\![1, m]\!]$, $\langle u_i, w\rangle \geqslant 0$.*

III.1. Le but de cette question est de montrer que C est un convexe fermé de $\mathbb{R}^n$.

(a) Montrer que C est convexe.

(b) Montrer que si $(u_1, \ldots, u_m)$ est une famille libre, alors C est fermé.

(c) Pour tout $\mathrm{I} \subset [\![1, m]\!]$, on pose $\mathrm{C}_\mathrm{I} = \left\{\sum_{i\in \mathrm{I}} \mu_i u_i,\ \forall i \in \mathrm{I}\ \mu_i \geqslant 0\right\}$. Montrer que

$$\mathrm{C} = \bigcup_{\mathrm{I}} \mathrm{C}_\mathrm{I}$$

où l'union est prise sur les ensembles $\mathrm{I} \subset [\![1, m]\!]$ tels que $(u_i)_{i\in\mathrm{I}}$ est une famille libre. En déduire que C est fermé.

III.2. On considère un vecteur $v \in \mathbb{R}^n \setminus \mathrm{C}$.

(a) Montrer que $\langle \mathrm{P}_\mathrm{C}(v), \mathrm{P}_\mathrm{C}(v) - v\rangle = 0$.

(b) On pose $w = \mathrm{P}_c(v) - v$. Montrer que $\langle v, w\rangle < 0$ et $\langle u_i, w\rangle \geqslant 0$ pour tout $i \in [\![1, m]\!]$.

III.3. Conclure la preuve du lemme 1.

Condition nécessaire d'optimalité

Soit $p \in \mathbb{N}^*$. Dans toute la suite de cette partie, on suppose que $f, g_1, \ldots, g_p$ sont des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ différentiables sur $\mathbb{R}^n$, et que

$$\mathrm{K} = \{x \in \mathbb{R}^n,\ g_1(x) \leqslant 0, \ldots, g_p(x) \leqslant 0\}$$

est non vide. Pour tout $x \in \mathrm{K}$, on note

$$\mathrm{I}_x = \{i \in [\![1, p]\!],\ g_i(x) = 0\}$$

III.4. Montrer que pour tout $x \in \mathrm{K}$,

$$\mathcal{A}_\mathrm{K}(x) \subset \{h \in \mathbb{R}^n,\ \forall i \in \mathrm{I}_x,\ \langle \nabla g_i(x), h\rangle \leqslant 0\}$$

III.5. On considère $x^* \in \mathrm{K}$ et on fait l'hypothèse suivante :

$$\exists v \in \mathbb{R}^n,\ \forall i \in \mathrm{I}_{x^*},\ \langle \nabla g_i(x^*), v\rangle < 0 \tag{H}$$

Montrer que $\mathcal{A}_\mathrm{K}(x^*) = \{h \in \mathbb{R}^n,\ \forall i \in \mathrm{I}_{x^*},\ \langle \nabla g_i(x^*), h\rangle \leqslant 0\}$.

III.6. Montrer que si $x^* \in \mathrm{K}$ est tel que $(\nabla g_i(x^*))_{i\in \mathrm{I}_{x^*}}$ forme une famille libre, alors l'hypothèse (H) est vérifiée.

III.7. On suppose que f atteint en $x^* \in \mathrm{K}$ un minimum local sur K, et que l'hypothèse (H) est vérifiée. Montrer qu'il existe des réels positifs $\mu_1^*, \ldots, \mu_p^*$ tels que

$$\begin{cases} \nabla f(x^*) + \sum\limits_{i=1}^{p} \mu_i^* \nabla g_i(x^*) = 0, \\ \forall i \in [\![1, p]\!],\ \mu_i^* g_i(x^*) = 0 \end{cases} \tag{1}$$

III.8. On suppose dans cette question que les fonctions $f, g_1, \ldots, g_p$ sont convexes. Soient $x^* \in \mathrm{K}$ et $\mu_1^*, \ldots, \mu_p^*$ tels que (1) soit vérifié. Montrer que f admet en x^* un minimum global sur K.

IV - Étude du problème dual

Le but de cette partie est d'aborder la minimisation d'une fonction f sur un sous-domaine K de $\mathbb{R}^n$ en considérant le problème « dual » associé. Dans un cas particulier, on propose une approche basée sur l'étude du problème dual pour obtenir une approximation du minimum de f sur K.

Soit $p \in [\![1, n]\!]$. On suppose dans toute cette partie que $f, g_1, \ldots, g_p$ sont des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ différentiables. On fait l'hypothèse supplémentaire que f est α-convexe pour un certain $\alpha \in \mathbb{R}_+^*$, et que les fonctions $g_1, \ldots, g_p$ sont convexes. On suppose par ailleurs que

$$\mathrm{K} = \{x \in \mathbb{R}^n,\ g_1(x) \leqslant 0, \ldots, g_p(x) \leqslant 0\}$$

est non vide. Dans toute la suite, on note $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{pmatrix}$ pour tout $x \in \mathbb{R}^n$.

On introduit la fonction $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^p \longrightarrow \mathbb{R}$ définie par

$$\mathcal{L}(x, \mu) = f(x) + \sum_{i=1}^{p} \mu_i g_i(x)$$

pour tout $x \in \mathbb{R}^n$ et tout $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}_+^p$. On s'intéresse au problème : trouver $x^* \in \mathrm{K}$ tel que

$$f(x^*) = \inf_{x \in \mathrm{K}} f(x) \tag{P}$$

IV.1. Montrer que

$$\inf_{x \in \mathrm{K}} f(x) = \inf_{x \in \mathbb{R}^n} \sup_{\mu \in \mathbb{R}_+^p} \mathcal{L}(x, \mu)$$

IV.2. Montrer que pour tout $\mu \in \mathbb{R}_+^p$, il existe un unique $x_\mu \in \mathbb{R}^n$ vérifiant $\mathcal{L}(x_\mu, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu)$.

Pour tout $\mu \in \mathbb{R}_+^p$, on note

$$\mathrm{G}(\mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu) = \mathcal{L}(x_\mu, \mu)$$

On va s'intéresser au problème dit dual : trouver $\mu^* \in \mathbb{R}_+^p$ tel que

$$\mathrm{G}(\mu^*) = \sup_{\mu \in \mathbb{R}_+^p} \mathrm{G}(\mu) = \sup_{\mu \in \mathbb{R}_+^p} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu) \tag{Q}$$

On dit que $(\overline{x}, \overline{\mu}) \in \mathbb{R}^n \times \mathbb{R}_+^p$ est un **point selle de $\mathcal{L}$** si

$$\mathcal{L}(\overline{x}, \overline{\mu}) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \overline{\mu}) \quad \text{et} \quad \mathcal{L}(\overline{x}, \overline{\mu}) = \sup_{\mu \in \mathbb{R}_+^p} \mathcal{L}(\overline{x}, \mu)$$

IV.3. On suppose dans cette question que $(\overline{x}, \overline{\mu}) \in \mathbb{R}^n \times \mathbb{R}_+^p$ est un point selle de $\mathcal{L}$.

(a) Montrer que $\overline{x}$ est solution de (P).

(b) Montrer que $\overline{\mu}$ est solution de (Q).

(c) Montrer que

$$\inf_{x \in \mathbb{R}^n} \sup_{\mu \in \mathbb{R}_+^p} \mathcal{L}(x, \mu) = \sup_{\mu \in \mathbb{R}_+^p} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu)$$

IV.4. On considère $x^* \in \mathrm{K}$ une solution de (P) satisfaisant l'hypothèse (H). Soit $\mu^* = (\mu_1^*, \ldots, \mu_p^*)$ comme dans la question III.7. Montrer que μ^* est solution de (Q).

IV.5. On suppose dans toute cette question que la fonction $\mu \in \mathbb{R}_+^p \mapsto x_\mu$ est continue. On considère une solution $\overline{\mu} \in \mathbb{R}_+^p$ de (Q).

(a) Soient $\mu \in \mathbb{R}_+^p$ et $\xi \in \mathbb{R}^p$ tels que $\mu + \xi \in \mathbb{R}_+^p$. Montrer que pour tout $t \in [0, 1]$, $\mu + t\xi \in \mathbb{R}_+^p$, et

$$\lim_{t \to 0^+} \frac{\mathrm{G}(\mu + t\xi) - \mathrm{G}(\mu)}{t} = \langle g(x_\mu), \xi \rangle$$

En déduire que pour tout $\mu \in \mathbb{R}_+^p$, $\langle g(x_{\overline{\mu}}), \mu - \overline{\mu} \rangle \leqslant 0$.

(b) Montrer que $x_{\overline{\mu}}$ est solution de (P).

IV.6. (Théorème d'Uzawa). Soient $\mathrm{A} \in \mathcal{M}_{p,n}(\mathbb{R})$ une matrice de rang p et $b \in \mathbb{R}^p$. On suppose que la fonction g est de la forme

$$g : x \mapsto \mathrm{A}x + b$$

(a) Montrer que pour tout $\mu \in \mathbb{R}_+^p$, $\nabla f(x_\mu) = -\mathrm{A}^\top \mu$, et en déduire que la fonction $\mu \mapsto x_\mu$ est continue sur $\mathbb{R}_+^p$.

(b) Montrer que (P) admet une unique solution $x^* \in \mathrm{K}$ et que (Q) admet une unique solution $\mu^* \in \mathbb{R}_+^p$.

Soit $\rho > 0$. On définit la suite $(\mu^k)_{k \in \mathbb{N}}$ par récurrence de la manière suivante :
- on fixe $\mu^0 \in \mathbb{R}_+^p$
- pour tout $k \in \mathbb{N}$, on pose $\mu^{k+1} = \mathrm{P}_{\mathbb{R}_+^p}(\mu^k + \rho g(x_{\mu^k}))$

où $\mathrm{P}_{\mathbb{R}_+^p} : \mathbb{R}^p \to \mathbb{R}_+^p$ est la projection sur le convexe fermé $\mathbb{R}_+^p$ de $\mathbb{R}^p$.

(a) Montrer que $\mu^* = \mathrm{P}_{\mathbb{R}_+^p}(\mu^* + \rho g(x_{\mu^*}))$.

On suppose désormais que $\|\mathrm{A}x\| \leqslant \sqrt{\frac{\alpha}{\rho}} \|x\|$ pour tout $x \in \mathbb{R}^n$.

(a) Montrer que la suite $(x_{\mu^k})_{k \in \mathbb{N}}$ converge vers x^*.

(b) Montrer que la suite $(\mu^k)_{k \in \mathbb{N}}$ converge vers μ^*.