SPÉCIALE MP* : DEVOIR SURVEILLÉ

La qualité de la rédaction et de la présentation, la clarté et la précision des raisonnements ainsi que la justification des calculs constitueront des éléments importants dans l'appréciation des copies.

Notations générales

$\mathbb{R}^n$ est muni de sa base canonique et du produit scalaire canonique noté $< .,. >$; la norme euclidienne sur $\mathbb{R}^n$ est notée $\|.\|$.

On identifie les applications linéaires de $\mathbb{R}^m$ dans $\mathbb{R}^n$ et les matrices de $\mathcal{M}_{n,m}(\mathbb{R})$ qui les représentent dans les bases canoniques ; de même on identifie naturellement les vecteurs de $\mathbb{R}^n$ et les matrices de $\mathcal{M}_{n,1}(\mathbb{R})$.

$\mathcal{S}_n(\mathbb{R})$: ensemble des matrices symétriques de $\mathcal{M}_n(\mathbb{R})$.

A^{T} est la notation utilisée pour la transposée de la matrice A.

$A \in \mathcal{M}_n(\mathbb{R})$ est dite symétrique semi-définie positive si, et seulement si,

$$A \in \mathcal{S}_n(\mathbb{R}) \text{ et } < Ax, x > \geqslant 0 \text{ pour tout } x \in \mathbb{R}^n.$$

$A \in \mathcal{M}_n(\mathbb{R})$ est dite symétrique définie positive si, et seulement si,

$$A \in \mathcal{S}_n(\mathbb{R}) \text{ et } < Ax, x > > 0 \text{ pour tout } x \text{ non nul dans } \mathbb{R}^n.$$

Pour une fonction numérique f définie sur un ouvert $\mathcal{O}$ de $\mathbb{R}^n$, différentiable en $x \in \mathcal{O}$, $\nabla f(x)$ désigne le (vecteur) gradient de f en x.

Présentation du problème

La première partie du problème concerne l'obtention des conditions d'optimalité du 1 $^\circ$ ordre pour un problème d'optimisation avec contraintes, par une méthode itérative de pénalisation.

La deuxième partie utilise les résultats de la première partie, on recherche la direction à choisir dans $\mathbb{R}^n$ dans un algorithme d'optimisation du second ordre (i.e. par exemple, si on cherche le minimum d'une fonction, dans quelle direction va-t-on le trouver si on part d'un point de $\mathbb{R}^n$?).

Préliminaires

Soit $f : \mathcal{O} \to \mathbb{R}$ où $\mathcal{O}$ est un ouvert convexe non vide de $\mathbb{R}^n$, on dit que f est convexe sur $\mathcal{O}$ ssi

$$\forall (a,b) \in \mathcal{O}^2, \forall t \in [0,1], f(ta + (1-t)b) \leqslant tf(a) + (1-t)f(b)$$

et que f est strictement convexe ssi

$$\forall (a,b) \in \mathcal{O}^2, a \neq b, \forall t \in]0,1[, f(ta + (1-t)b) < tf(a) + (1-t)f(b).$$

On rappelle les propriétés vraies pour les fonctions de classe $\mathcal{C}^1$ d'une variable ($\mathcal{O}$ est ici un intervalle ouvert de $\mathbb{R}$) :

(i) f est convexe sur $\mathcal{O}$ ssi $\forall (x,y) \in \mathcal{O}^2, f(y) \geqslant f(x) + f'(x)(y-x)$,

(ii) pour une fonction f de classe $\mathcal{C}^2$,

$$f \text{ est convexe } \Leftrightarrow \forall x \in \mathcal{O}, f''(x) \geqslant 0$$

$$(\forall x \in \mathcal{O}, f''(x) > 0) \Rightarrow f \text{ est strictement convexe.}$$

$\boxed{0^\circ}$

Prouver alors les propriétés suivantes :

<u>0.1</u> Montrer que, si f est une fonction de classe $\mathcal{C}^1$ sur $\mathcal{O}$ ouvert convexe non vide de $\mathbb{R}^n$ alors, on a l'équivalence :

$$f \text{ convexe } \Leftrightarrow (\forall (x,y) \in \mathcal{O}^2,\ f(y) \geqslant f(x) + f'(x)(y-x) = f(x) + < \nabla f(x), y - x > .$$

0.2 Montrer que, si f est convexe, de classe $\mathcal{C}^1$ sur $\mathcal{O}$ ouvert convexe non vide de $\mathbb{R}^n$ alors on a l'équivalence suivante :

$$f \text{ présente un minimum en } x^* \Leftrightarrow \nabla f(x^*) = 0.$$

0.3 On suppose maintenant que f est de classe $\mathcal{C}^2$ sur $\mathcal{O}$ ouvert convexe non vide, on définit, pour tout $x \in \mathcal{O}$, $A(x)$ comme étant la matrice $\left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)$.

Prouver que (en se limitant au cas où $n = 2$) :

$$f \text{ est convexe } \Leftrightarrow \forall x \in \mathcal{O}, A(x) \text{ est symétrique positive}$$

et que

$$(\forall x \in \mathcal{O}, A(x) \text{ définie positive }) \Rightarrow f \text{ est strictement convexe.}$$

Partie I

Soit $\mathcal{O}$ un ouvert de $\mathbb{R}^n$, $f\ : \mathcal{O} \to \mathbb{R}$ une fonction continûment différentiable sur $\mathcal{O}$, $g_1, g_2, \dots, g_p$ ($p \in \mathbb{N}^*$), $h_1, h_2, \dots, h_q$ ($q \in \mathbb{N}^*$) des fonctions continûment différentiables de $\mathbb{R}^n$ dans $\mathbb{R}$.

L'ensemble S des contraintes est défini comme suit :

$$S = \{x \in \mathbb{R}^n |\ g_i(x) \leqslant 0 \text{ pour tout } i = 1, \dots, p \text{ et } h_j(x) = 0 \text{ pour tout } j = 1, \dots, q\}$$

(contraintes sous forme d'*inégalités* et sous forme d'*égalités* ; toutefois les situations étudiées dans le problème ne feront pas apparaître nécessairement les deux formes de contraintes). On suppose que S est non vide et contenu dans $\mathcal{O}$.

$x^* \in S$ est dit *minimum local* de f sur S s'il existe $r > 0$ tel que $\overline{B}(x^*, r) \subset \mathcal{O}$ et $f(x) \geqslant f(x^*)$ pour tout $x \in S \cap \overline{B}(x^*, r)$.

On dit que $x^* \in S$ est un *minimum global* de f sur S (ou que x^* *minimise* f *sur* S) si $f(x) \geqslant f(x^*)$ pour tout $x \in S$.

$\boxed{1^\circ}$ Soit $x^* \in S$ un minimum local de f sur S, minimisant f sur $S \cap \overline{B}(x^*, r)$.

Pour tout $k \in \mathbb{N}^*$ on définit

$$\begin{array}{rcl} \Phi_k\ : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q & \to & \mathbb{R} \\ U = (u, v_1, \dots, v_p, w_1, \dots, w_q) & \mapsto & \Phi_k(U) = u + k \sum\limits_{i=1}^{p} (v_i^+)^2 + k \sum\limits_{j=1}^{q} w_j^2 \end{array}$$

(v_i^+ désigne ici $\max\{0, v_i\}$).

et $F_k\ : \mathcal{O} \to \mathbb{R}$ par :

$$\begin{aligned} x \in \mathcal{O} \mapsto F_k(x) &= \Phi_k(f(x), g_1(x), \dots, g_p(x), h_1(x), \dots, h_q(x)) + \|x - x^*\|^2 \\ &= f(x) + k \sum_{i=1}^{p} \left[g_i^+(x)\right]^2 + k \sum_{j=1}^{q} [h_j(x)]^2 + \|x - x^*\|^2 \end{aligned}$$

On désigne par x_k un point de $\overline{B}(x^*, r)$ minimisant F_k sur $\overline{B}(x^*, r)$.

1.1 Prouver que pour tout $k \in \mathbb{N}^*$, $F_k(x_k) \leqslant f(x^*)$.

En déduire :

(i) : $\lim\limits_{k \to +\infty} g_i^+(x_k) = 0$ pour tout $i = 1, \dots, p$

et

(ii) : $\lim\limits_{k \to +\infty} h_j(x_k) = 0$ pour tout $j = 1, \dots, q$.

1.2 Prouver que si $\tilde{x}$ est la limite d'une suite extraite de la suite $(x_k)_{k \in \mathbb{N}^*}$, alors $\tilde{x} = x^*$.

En déduire que la suite $(x_k)_{k \in \mathbb{N}^*}$ converge vers x^* quand $k \to +\infty$.

1.3

1.3.1 Calculer $\nabla\Phi_k(U)$, sa norme euclidienne $\|\nabla\Phi_k(U)\|$ (on pourra commencer par calculer la *dérivée* de $x \mapsto (x^+)^2$).

En posant $U_k = (f(x_k), g_1(x_k), \dots, g_p(x_k), h_1(x_k), \dots, h_q(x_k))$ montrer que l'on peut extraire de la suite

$$\left(\frac{\nabla\Phi_k(U_k)}{\|\nabla\Phi_k(U_k)\|}\right)_{k\in\mathbb{N}^*}$$

une suite qui converge vers un vecteur

$$(\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q)$$

vérifiant

(α) $\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ ne sont pas tous nuls ;
(β) $\lambda_i \geqslant 0$ pour tout $i = 0, 1, \dots, p$;
1.3.2 Que vaut $\nabla F_k(x_k)$? Montrer encore que :
(γ) $\lambda_i g_i(x^*) = 0$ pour tout $i = 1, \dots, p$;
(δ) $\lambda_0 \nabla f(x^*) + \sum\limits_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum\limits_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0.$

On dira que $\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ sont des *multiplicateurs positivement homogènes* en x lorsque les conditions (α), (β), (γ) et (δ) précédentes sont satisfaites en x.

$\boxed{2^\circ}$ Dans cette question on fait l'hypothèse suivante (sur x élément générique de S) :

$(\mathbf{QC})_x$ Les vecteurs $\nabla h_1(x), \dots, \nabla h_q(x)$ sont linéairement indépendants et il existe $d \in \mathbb{R}^n$ vérifiant :

$< \nabla g_i(x), d > < 0$ pour tout i tel que $g_i(x) = 0$

et

$< \nabla h_j(x), d >= 0$ pour tout $j = 1, \dots, q$.

2.1 Montrer que si $\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ sont des multiplicateurs positivement homogènes en x, alors $\lambda_0 > 0$ nécessairement.

2.2

On dit que $\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ sont des *multiplicateurs d'ordre un* en x lorsque

$1, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q$ sont des multiplicateurs positivement homogènes en x.

L'ensemble des multiplicateurs d'ordre 1 en x est noté $\mathcal{M}_1(x)$. On suppose $\mathcal{M}_1(x)$ non vide.

Montrer que $\mathcal{M}_1(x)$ est un convexe compact non vide de $\mathbb{R}^{p+q}$ (une illustration géométrique des propriétés des multiplicateurs d'ordre 1 peut rendre service).

$\boxed{3^\circ}$ On fait pour cette question les hypothèses suivantes sur les données :

$\mathcal{O}$ est convexe, $f\ : \mathcal{O} \to \mathbb{R}$ est convexe et de classe $\mathcal{C}^1$ sur $\mathcal{O}$;
$g_i\ : \mathbb{R}^n \to \mathbb{R}$ est convexe et de classe $\mathcal{C}^1$ pour tout $i = 1, \dots, p$;
$h_j\ : \mathbb{R}^n \to \mathbb{R}$ est affine pour tout $j = 1, \dots, q$, c'est-à-dire de la forme
$x \mapsto h_j(x) =< a_j, x > -b_j$ avec $a_j \in \mathbb{R}^n$ et $b_j \in \mathbb{R}$.

3.1 On considère la condition suivante :

$(\mathbf{QCC})$ Les vecteurs $a_1, \dots, a_q$ sont linéairement indépendants et il existe $\bar{x} \in \mathbb{R}^n$ vérifiant :

$g_i(\bar{x}) < 0$ pour tout $i = 1, \dots, p$

et

$< a_j, \bar{x} > -b_j = 0$ pour tout $j = 1, \dots, q$.

Montrer que $(\mathbf{QCC})$ implique $(\mathbf{QC})_x$ pour tout point x de S.

<u>3.2</u> On suppose que (**QCC**) est vérifiée. Montrer l'équivalence des énoncés suivants concernant $x \in S$:

(e_1) x est un minimum global de f sur S ;

(e_2) x est un minimum local de f sur S ;

(e_3) il existe des multiplicateurs d'ordre un en x ;

(on pourra considérer la fonction $\mathcal{L}\ : u \in \mathcal{O} \mapsto f(u) + \sum\limits_{i=1}^{p} \lambda_i g_i(u) + \sum\limits_{j=1}^{q} \mu_j h_j(u)\big)$.

$\boxed{4^\circ}$ On traite ici un exemple de problème de minimisation (avec contraintes) dans $\mathbb{R}^n$. Les données du problème sont :

$a_1, \ldots, a_n, b_1, \ldots, b_n, \sigma$ des réels (fixés) tous strictement positifs ;

$x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mapsto f(x) = \sum\limits_{i=1}^{n} a_i \left(e^{-b_i x_i} - 1\right)$;

$S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \quad | \quad \sum\limits_{i=1}^{n} x_i = \sigma$ et $x_i \geqslant 0$ pour tout $i = 1, \ldots, n\}$.

<u>4.1</u> Montrer qu'il existe un et un seul $x^* = (x_1^*, \ldots, x_n^*) \in S$ minimisant (globalement) f sur S.

<u>4.2</u> Indiquer pourquoi l'équation

$$\Phi(r) = \sum_{\{i | r < a_i b_i\}} \frac{1}{b_i} \ln \frac{a_i b_i}{r} = \sigma, \quad r > 0$$

a une seule solution, que l'on notera r^*

(ln désigne ici le logarithme népérien et $\sum\limits_{\emptyset} = 0$ par convention).

<u>4.3</u> Exprimer les coordonnées x_i^* du minimum global x^* de f sur S en fonction de r^* et $a_1, \ldots, a_n, b_1, \ldots, b_n$.

Partie II

<u>Notations</u>

$\mathbb{R}_+$: ensemble des réels positifs.

Étant donné $A \in \mathcal{M}_n(\mathbb{R})$ symétrique semi-définie positive, on appelle pseudo-inverse de A, et on note A^+, l'application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$ définie comme suit :

Pour $y \in \mathbb{R}^n$, soit $\bar{y}$ la projection orthogonale de y sur $\operatorname{Im} A$; il existe alors un et un seul $x \in \operatorname{Im} A$ tel que $Ax = \bar{y}$. L'application (linéaire) $y \mapsto x = A^+ y$ est la pseudo-inverse de A.

Étant donné $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, on appelle factorisation de Cholesky de A la factorisation $A = CC^{\mathrm{T}}$, où $C \in \mathcal{M}_n(\mathbb{R})$ est triangulaire inférieure avec tous ses éléments diagonaux strictement positifs.

<u>Présentation</u> : Soit $\delta > 0$, $b \in \mathbb{R}^n$ et $B \in \mathcal{S}_n(\mathbb{R})$. On considère le problème de minimisation suivant :

$(\mathcal{P})$ Minimiser $f\ : x \mapsto f(x) = \dfrac{1}{2} < Bx, x > + < b, x >$ sur $S = \overline{B}(0, \delta)$.

On dira que $x^* \in S$ est solution de $(\mathcal{P})$ lorsque x^* est un minimum global de f sur S. Résoudre $(\mathcal{P})$ signifie donc trouver les minima globaux de f sur S.

$\boxed{5^\circ}$ On commence par traiter un exemple simple : (*et qui sera révélateur pour la suite*)

$n = 2$, $b = (1, 0)$, $B = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}$, où α est un réel. Trouver la (ou les) solution(s) de $(\mathcal{P})$ suivant δ et α.

Il est fortement conseillé d'utiliser les résultats de la première partie et de distinguer les cas : $\alpha > 0$, $\alpha = 0$ et $\alpha < 0$.

$\boxed{6^\circ}$

6.1 On traite ici des cas où des points intérieurs à S pourraient être solutions de $(\mathcal{P})$.

6.1.1 Montrer que

$$x^* = 0 \text{ est solution de } (\mathcal{P}) \Leftrightarrow b = 0, B \text{ semi-définie positive.}$$

6.1.2 Montrer que :
il n'y a pas de solutions de $(\mathcal{P})$ sur la frontière de S ssi B est définie positive et $\|B^{-1}b\| < \delta$.
On pourra prouver et utiliser la formule suivante :

$$f(x+h) = f(x) + < \nabla f(x), h > + \frac{1}{2} < \nabla^2 f(x)h, h > .$$

6.2

6.2.1 On suppose que x^*, solution de $(\mathcal{P})$, vérifie $\|x^*\| = \delta$.
Montrer qu'il existe $\lambda \geqslant 0$ tel que :

(α) : $(B + \lambda I_n)x^* + b = 0$;
(β) : $B + \lambda I_n$ est semi-définie positive.
Utiliser les résultats de la question n ° 2

6.2.2 Réciproquement, soit x^* vérifiant $\|x^*\| = \delta$ et pour lequel il existe $\lambda \geqslant 0$ assurant (α) et (β). Montrer que x^* est alors solution de $(\mathcal{P})$.

6.2.3 Soit x^* vérifiant $\|x^*\| = \delta$ et pour lequel il existe $\lambda \geqslant 0$ assurant (α) et
(β') : $B + \lambda I_n$ est définie positive.
Vérifier que x^* est alors la seule solution de $(\mathcal{P})$.

6.3 Rassembler tous les cas traités en un énoncé clair et précis donnant une condition nécessaire et suffisante pour que $x^* \in S$ soit un minimum global de f sur S.

$\boxed{7^\circ}$

7.1 Résoudre le problème $(\mathcal{P})$ lorsque $B = 0$ et $b \neq 0$.

Dans toute la suite on désignera par λ_1 la plus petite valeur propre de B et par E_1 le sous-espace propre associé.

7.2 Résoudre le problème $(\mathcal{P})$ lorsque $b = 0$.

On suppose désormais $B \neq 0$ et $b \neq 0$. Pour tout $\lambda > -\lambda_1$ on désigne par $x(\lambda)$ la seule solution de

$$(B + \lambda I_n)x + b = 0,$$

et on pose $\theta(\lambda) = \|x(\lambda)\|$.

7.3

7.3.1 Montrer que l'on peut écrire

$$x(\lambda) = -\sum_{i=1}^{n} \frac{< b, v_i >}{\lambda_i + \lambda} v_i$$

où les λ_i sont les valeurs propres de B et où les v_i sont des vecteurs à préciser.

7.3.2 Étudier les variations et la convexité de la fonction θ sur $]-\lambda_1, +\infty[$. On s'attachera notamment à déterminer $\lim_{\lambda \downarrow -\lambda_1} \theta(\lambda)$ dans les différents cas possibles.

7.4

7.4.1 Déterminer $(B - \lambda_1 I_n)^+$.

7.4.2 A quelles conditions l'équation $\theta(\lambda) = \delta$ a t-elle des solutions sur $] - \lambda_1, +\infty[\cap \mathbb{R}_+$? Dans le cas où $\lambda_1 < 0$, comment ces racines conduisent-elles aux solutions de $(\mathcal{P})$?

7.5 On définit la fonction $\psi\ :]-\lambda_1,+\infty[\to\mathbb{R}$ comme suit :

$$\lambda\in]-\lambda_1,+\infty[\mapsto\psi(\lambda)=\frac{1}{\delta}-\frac{1}{\theta(\lambda)}.$$

Montrer que ψ est convexe.

7.6 Dans cette question et la suivante 7.7, on fait l'hypothèse supplémentaire suivante :

$$\lambda_1<0,\ b \text{ n'est pas orthogonal à } E_1 \text{ ou bien } \|(B-\lambda_1 I_n)^+b\|>\delta.$$

On considère l'algorithme suivant :

- Étape 0 : initialisation, $\lambda^0\in]-\lambda_1,+\infty[$.
- Étape k : construction de λ^{k+1} à partir de λ^k :

Soit $\lambda\mapsto\bar\theta_k(\lambda)=\dfrac{a_k}{b_k+\lambda}-\delta$ la fonction rationnelle de λ telle que

$$\bar\theta_k(\lambda^k)=\theta(\lambda^k),\quad \bar\theta_k'(\lambda^k)=\theta'(\lambda^k).$$

On prend pour λ^{k+1} la racine de l'équation $\bar\theta_k(\lambda)=0$.

7.6.1 Exprimer λ^{k+1} en fonction de $\lambda^k,\theta(\lambda^k),\theta'(\lambda^k)$ et δ.

7.6.2 Comparer l'algorithme précédent à l'algorithme de Newton pour résoudre l'équation $\psi(\lambda)=0$.

7.7 On considère l'algorithme suivant :

- Étape 0 : initialisation, $\lambda^0\in]-\lambda_1,+\infty[$ avec $\theta(\lambda^0)>\delta$.
- Étape k : construction de λ^{k+1} à partir de λ^k :
  . Factorisation CC^{T} de Cholesky de $B+\lambda^k I_n$;
  . Résolution de $CC^{\mathrm{T}}u=-b$;
  . Résolution de $Cv=u$;

$$\lambda^{k+1}=\lambda^k+\left(\frac{\|u\|}{\|v\|}\right)^2\left(\frac{\|u\|-\delta}{\delta}\right).$$

Vérifier que cet algorithme est l'algorithme de Newton pour résoudre $\psi(\lambda)=0$.

Quels résultats de convergence peut-on donner à son sujet ?