INSA Toulouse - Département GMM **Jeudi 19 janvier 2012**
4ème année

CC2 - Optimisation

Durée : 2h.

Seuls le polycopié de cours et les notes personnelles de cours sont autorisés.

Partie optimisation

Exercice 1. Quelques questions élémentaires

1. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction fortement convexe. Montrez qu'elle admet un minimiseur unique.

   Solution : f fortement convexe $\Rightarrow$ f strictement convexe $\Rightarrow$ f admet un minimiseur unique.

2. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction μ-fortement convexe. On note x^* son minimiseur. Montrer que si $f(x) - f(x^*) \leq \epsilon$ alors $\|x - x^*\|_2^2 \leq C\epsilon$. Préciser la valeur de C.

   Solution : On a :

$$f(y) \geq f(x) + \langle \eta, y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2,\ \forall (y, x) \in \mathbb{R}^n \times \mathbb{R}^n,\ \forall \eta \in \partial f(x). \quad (1)$$

   Donc en particulier pour $x = x^*$ comme $0 \in \partial f(x^*)$:

$$f(y) \geq f(x^*) + \frac{\mu}{2}\|y - x^*\|_2^2,\ \forall y \in \mathbb{R}^n \quad (2)$$

   et donc :

$$\|y - x^*\|_2^2 \leq \frac{2}{\mu}(f(y) - f(x^*)),\ \forall y \in \mathbb{R}^n. \quad (3)$$

   D'où $C = \frac{2}{\mu}$.

3. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction définie par $f(x) = \|Ax - b\|_p^p$ [1] avec $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ et $p \in \mathbb{R}_+^*$. Dire, en le justifiant de manière concise, pour quelles valeurs de p la fonction f est convexe. Pour quelles valeurs de p est-elle différentiable ? Pour chaque cas, proposer un algorithme de minimisation.

   Solution : Pour $p \geq 1$, $\|y\|_p^p$ est une norme (donc une fonction convexe) composée avec une fonction croissante convexe (la puissance p). f est donc convexe. Pour $p < 1$, $\|y\|_p^p$ n'est pas une norme. La hessienne de $y \mapsto \|y\|_p^p$ se calcule facilement (elle est diagonale) et ses valeurs propres sont négatives. La fonction f n'est donc pas convexe pour $p < 1$.

   La fonction est différentiable pour $p > 1$. Pour $p \leq 1$, la fonction 1D $x \mapsto x^p$ n'est pas dérivable en 0, donc f n'est pas dérivable.

   Pour $p < 1$ aucun algorithme de minimisation n'a été vu en cours. Pour $p = 1$ on peut utiliser un algorithme de descente de sous-gradient, un algorithme de point intérieurs ou un algorithme du simplexe. Pour $p > 1$, les descentes de gradient peuvent être utilisées.

---

1. On rappelle que $\|x\|_p = \left(\sum_{i=1}^n x_i^p\right)^{1/p}$ pour tout $p > 0$.

4. On considère la fonction $g : \mathbb{R}^m \to \mathbb{R}$ définie par $g(y) = \frac{1}{2}\||y| - b\|_2^2$ où $|y|$ est le vecteur dont la i-ème coordonnée est $|y_i|$. On pose $f(x) = g(Ax)$ où $A \in \mathbb{R}^{m\times n}$.

   (a) On considère le cas $m = 1$. Est-ce que g est convexe ?

   Solution : En dehors du cas $b = 0$, g n'est pas convexe. En 1D par exemple, pour $b = 1$, cette fonction admet deux minima en -1 et 1. Or l'ensemble $\{-1, 1\}$ n'est pas convexe et une fonction convexe admet un ensemble convexe ou vide de minimiseurs.

   (b) On considère le cas $m = 1$. Montrez que g n'est pas dérivable en 0 pour $b \neq 0$.

   Solution :

   Pour $b = 0$, la fonction est convexe, c'est une simple fonction quadratique. On choisit $b \neq 0$. On a

$$g(0+h) = \begin{cases} \frac{1}{2}(h-b)^2 & \text{si } h \geq 0 \\ \frac{1}{2}(-h-b)^2 & \text{sinon} \end{cases} \tag{4}$$

   soit encore :

$$g(0+h) = \begin{cases} g(0) - hb + o(h) & \text{si } h \geq 0 \\ g(0) + hb + o(h) & \text{sinon} \end{cases} \tag{5}$$

   or l'application :

$$h \mapsto \begin{cases} -2hb & \text{si } h \geq 0 \\ 2hb & \text{sinon} \end{cases} \tag{6}$$

   n'est pas linéaire.

   (c) Calculer ∇g pour $y \in Y$ où $Y = \{y \in \mathbb{R}^m, y_i \neq 0, \forall i \in \{1, \cdots, m\}\}$.

   Solution :

   La fonction g s'écrit $g(y) = \sum_{i=1}^m \frac{1}{2} g_i(y_i)$ où $g_i(y_i) = \frac{1}{2}(|y_i| - b_i)^2$. Donc :

$$\nabla g(y) = \begin{pmatrix} g_1'(y_1) \\ \vdots \\ g_m'(y_m) \end{pmatrix}$$

   avec

$$g_i'(y_i) = \begin{cases} (y_i - b_i) & \text{si } y_i > 0 \\ y_i + b_i & \text{si } y_i < 0. \end{cases} \tag{7}$$

   (d) Calculer ∇f sur l'ensemble $X = \{x \in \mathbb{R}^n, Ax \in Y\}$.

   Solution :

   On a $\nabla f(x) = A^T \nabla g(Ax)$.

   (e) Proposez un algorithme de minimisation de f.

   Solution :

   En réalité, on n'a vu aucun algorithme de minimisation pour des fonctions non différentiables et non convexes. Cette remarque est suffisante pour obtenir les points de la question.

   Une solution possible consiste à régulariser g, par exemple en posant :

$$g_\epsilon(y) = \frac{1}{2}\|\sqrt{y^2+\epsilon^2} - b\|_2^2.$$

Ainsi la fonction devient différentiable et on peut utiliser une méthode de minimisation locale de type descente de gradient. Des méthodes plus avancées de relaxation existent. Par exemple les relaxations semi-définies. C'est un problème rencontré fréquemment dans l'industrie.

Exercice 2. Cône normal et conditions d'optimalité

Soit $X \subseteq \mathbb{R}^n$ un ensemble fermé. On appelle cône normal au point $x \in \partial X$ [2] l'ensemble :

$$N_X(x) = \{d \in \mathbb{R}^n, \langle d, y - x\rangle \leq 0, \forall y \in X\}. \tag{8}$$

1. On considère les ensembles $X_1 = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ et $X_2 = \{(x, y) \in \mathbb{R}^2, 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Dessinez ces ensembles (sur des figures séparées) ainsi que les cônes normaux $N_{X_1}(1, 0)$, $N_{X_2}(0, 0)$ et $N_{X_2}(0.5, 1)$.
   Solution :
   Trop compliqué en Latex. Pour X_1, le cône normal est la normale extérieure au cercle unité. Pour X_2, $N_{X_2}(0.5, 1)$ est la normale éxterieure au bord du carré et $N_{X_1}(0, 0)$ est l'othant négatif.
2. Montrez que pour tout ensemble X fermé, $N_X(x)$ est un ensemble convexe.
   Solution :
   Immédiat. Vérifier que pour $(d_1, d_2) \in N_X(x) \times N_X(x)$, $\frac{d_1+d_2}{2} \in N_X(x)$.
3. Désormais X représente un sous-ensemble *convexe*, fermé de $\mathbb{R}^n$. On considère la fonction indicatrice de X définie par

$$h(x) = \chi_X(x) = \begin{cases} 0 & \text{si } x \in X \\ +\infty & \text{sinon.} \end{cases}$$

   Calculez le sous-différentiel de h en un point $x \in \overset{\circ}{X}$.
   Solution :
   Pour $x \in \overset{\circ}{X}$, h est nulle sur une boule suffisamment petit autour de x. Donc h est différentiable en x et son gradient est nul. Par conséquent $\partial h(x) = \{0\}$.
4. Montrez que $\partial h(x) = N_X(x)$ pour $x \in \partial X$.
   Solution :
   Pour $x \in \partial X$, on a :

$$\begin{aligned} \partial h(x) &= \{\eta \in \mathbb{R}^n, h(y) \geq h(x) + \langle \eta, y - x\rangle, \forall y \in \mathbb{R}^n\} \\ &= \{\eta \in \mathbb{R}^n, 0 \geq 0 + \langle \eta, y - x\rangle, \forall y \in X\} \\ &= \{\eta \in \mathbb{R}^n, 0 \geq \langle \eta, y - x\rangle, \forall y \in X\} \\ &= N_X(x). \end{aligned}$$

5. On considère maintenant la fonction $f(x) = g(x)+h(x)$ où $g : \mathbb{R}^n \to \mathbb{R}$ est une fonction convexe différentiable et h est la fonction définie dans la question précédente. Déterminez les conditions d'optimalité du problème non contraint :

$$\min_{x \in \mathbb{R}^n} f(x). \tag{9}$$

2. ∂X est la frontière de X définie comme $\partial X = \bar{S} \setminus \overset{\circ}{S}$

Solution :
Les conditions d'optimalité sont $0 \in \partial f(x)$ soit encore :

$$\nabla g(x) = 0 \text{ et } x \in \overset{\circ}{X}$$

ou

$$0 \in \nabla g(x) + N_X(x) \text{ et } x \in \partial X.$$

6. A quel problème d'optimisation contraint, (9) est-il équivalent ?
Solution :
Au problème :

$$\min_{x \in X} g(x).$$

7. On se place dans R^2, et on pose $h(x, y) = \chi_{X_2}(x, y)$ et $g(x, y) = \frac{1}{2}\left((x-2)^2 + (y-2)^2\right)$. Dessinez sur un même schéma quelques lignes de niveau de g ainsi que l'ensemble X_2. Trouvez graphiquement et justifiez en utilisant les résultats précédents où se trouve le minimiseur de la fonction $f = g + h$.
Solution :
Le minimiseur est unique (par forte convexité de g). Il se trouve au point $(1, 1)$. Il suffit de vérifier que $\nabla g(1, 1)$ appartient à l'orthant positif.

Partie algèbre linéaire

Note : cette partie sera rédigée sur une feuille différente de la partie précédente.

Exercice 3. Soient M et K deux matrices symétriques définies positives de $M_n(\mathbb{R})$. On s'intéresse au problème généralisé de valeurs propres

$$\lambda\, M\, x - K\, x = 0. \tag{10}$$

Montrez qu'il existe n valeurs propres λ_i, $1 \leq i \leq n$ et une base de vecteurs propres notés $(u^i)_{1\leq i\leq n}$ tels que
- $u^{i\ T}\, M\, u^j = u^{i\ T}\, K\, u^j = 0$ is $i \neq j$,
- $u^{i\ T}\, M\, u^i = 1$, $1 \leq i \leq n$,
- $u^{i\ T}\, K\, u^i = \lambda_i$, $1 \leq i \leq n$.

Exercice 4. Soit A une matrice symétrtique de $M_n(\mathbb{R})$, dont la valeur propre dominante unique notée λ_1 est simple, et P une matrice de diagonalisation de A

On applique à cette matrice la méthode de la puissance utilisant la norme euclidienne (Exemple 3.3 page 61 du polcopié). On note l_k l'approximation de λ_1 à l'étape k, q_k celle du vecteur propre unitaire associé, et $r_k = A\, q_k - q_k\, l_k l_k$ le résidu associé.

1. Montrez que $(l_k,\, q_k)$ est un élément propre de la matrice perturbée $A - E_k$, avec $E_k = r_k\, q_k^*$.
2. En déduire que
$$|l_k - \lambda_1| \leq \|r_k\|_2\, \mathcal{K}_2(P).$$
3. Préciser cette estimation en tenant compte de la symétrie de A.