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.

Exercice 1. Applications directes du cours (7 points)

1. Déterminez le gradient et la hessienne H_F de la fonction :

$$\begin{array}{rccc} F: & \mathbb{R}^n & \to & \mathbb{R} \\ & x & \mapsto & \frac{1}{2}\|Ax-b\|_2^2 \end{array} \tag{1}$$

   où $\|\cdot\|_2^2$ est la norme l^2 usuelle, $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^n$.
   Solution : On a $\nabla F(x) = A^T(Ax-b)$ et $H_F(x) = A^TA$.

2. A quelle condition sur la matrice A, la fonction F est-elle convexe ?
   Solution : Elle l'est toujours.

3. A quelle condition sur A est elle fortement convexe ?
   Solution : elle est fortement convexe si A est de rang plein. Ainsi A^TA est symétrique, définie positive et $\lambda_{min}(H_F(x)) = \lambda_{min}(A^TA) > 0, \ \forall x \in \mathbb{R}^n$.

4. Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction C^2, convexe et :

$$\begin{array}{rccc} \Phi: & \mathbb{R}^m & \to & \mathbb{R} \\ & y & \mapsto & \sum_{i=1}^m \phi(y_i) \end{array} \tag{2}$$

   Déterminer le gradient de la fonction $G(x) = \Phi(Ax-b)$.
   Solution : On a $\nabla G(x) = A^T\nabla\Phi(Ax-b)$ où :

$$\nabla\Phi(y) = (\phi'(y_1), \phi'(y_2), ..., \phi'(y_m)). \tag{3}$$

5. On souhaite minimiser F. Quel algorithme utiliseriez-vous si $n = 10$? Si $n = 10^6$?
   Solution : Un gradient conjugué linéaire est probablement la meilleure méthode sans connaissance supplémentaire sur A. Une méthode de descente de gradient accélérée serait aussi optimale en un certain sens. Ces remarques sont vraies si $n = 10$ ou $n = 10^6$. Si on connait plus d'informations sur A, il pourrait être bénéfique de préconditionner le problème.

6. On souhaite minimiser G. Quel algorithme utiliseriez-vous si $n = 10$? Si $n = 10^6$?
   Solution : Dans les deux cas, on peut utiliser une descente de gradient accélérée. C'est a priori une méthode optimale pour ce problème. On peut aussi la préconditionner si la structure de A est connue.

7. Quel algorithme utiliseriez-vous si $\Phi(y) = \|y\|_1$ si $n = 10$? si $n = 10^6$? si $n = 10^{19}$?
   Solution :
   C'est un problème de programmation linéaire. Pour $n = 10$ et $n = 10^6$, on peut donc utiliser l'algorithme du simplexe ou l'algorithme des points intérieurs. L'algorithme des points intérieurs a un bien meilleur comportement au pire des cas. Si $n = 10^{19}$, il est a priori impossible de résoudre (et même de poser) le problème actuellement. En effet, un tableau de 10^{19} nombres au format double fait plus de 160 giga - gigaoctets...

Exercice 2. Dualité de Fenchel-Rockafellar (14 points)

Soit f une fonction convexe fermée. On considère la fonction f^* définie de la façon suivante :

$$f^*(x) = \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f(y).$$

L'application $f \mapsto f^*$ est appelée *transformée de Fenchel-Rockafellar* ou *polaire* de f et c'est un outil fondamental d'analyse convexe. L'objectif de cet exercice est de déterminer quelques-unes de ses propriétés.

PARTIE 1 - Quelques exemples de polaires :

On pose $X = \{x \in \mathbb{R}^n, \|x\|_2 \leq 1\}$. On considère les fonctions :

$$f_1(x) = \frac{1}{2}\|x\|_2^2$$

$$f_2(x) = \chi_X(x) = \begin{cases} 0 & \text{si } \|x\|_2 \leq 1 \\ +\infty & \text{sinon} \end{cases}$$

$$f_3(x) = f_1(x) + f_2(x).$$

1. Chacune des fonctions ci-dessus est-elle convexe ? Fermée ?
   Solution : toutes ces fonctions sont convexes fermées. En effet, f_1 est convexe et continue. Elle est donc fermée. De plus X est un ensemble convexe fermé, donc son indicatrice est convexe, fermée. Enfin, f_3 est la somme de deux fonctions convexe, fermées, elle est donc convexe fermée.
2. Déterminez la transformée de Fenchel f_1^* de f_1. Que vaut f_1^{**} ?
   Solution : En annulant le gradient de la fonction $y \mapsto \langle x, y \rangle - \frac{1}{2}\|y\|_2^2$, on trouve qu'à l'optimum, $y = x$. En réinjectant cette quantité, on obtient

   $$f_1^*(x) = \frac{1}{2}\|x\|_2^2.$$

   Pour calculer $f_1^{**}(x)$, il suffit de se rendre compte que $f_1^*(x) = f_1(x)$. Donc $f_1^{**}(x) = (f_1^*)^*(x) = f_1^*(x) = f_1(x) = \frac{1}{2}\|x\|_2^2$.
3. Montrez que :

   $$f_2^*(x) = \sup_{y \in \mathbb{R}^n, \|y\|_2^2 \leq 1} \langle x, y \rangle.$$

   Solution : évident.

4. Calculez f_2^* et f_2^{**}.
Solution : On peut procéder de deux façons. Soit on utilise la remarque précédente et on utilise la dualité lagrangienne. Soit on remarque que d'après le théorème de Cauchy-Schwarz :
$$f_2^*(x) = \sup_{y \in \mathbb{R}^n, \|y\|_2 \leq 1} \langle x, y \rangle = \|x\|_2.$$
Pour le calcul de f_2^{**}, on a encore plusieurs choix.
– Soit on écrit :
$$f_2^{**}(x) = \sup_{y \in \mathbb{R}^n} \langle y, x \rangle - \|y\|_2.$$
En posant $J(y) = \langle y, x \rangle - \|y\|_2$, les conditions d'optimalité du problème deviennent $0 \in \partial J(y)$. On écrit alors le sous-différentiel de la norme l^2 pour conclure.
– Soit on écrit :
$$\begin{aligned} f_2^{**}(x) &= \sup_{y \in \mathbb{R}^n} \langle y, x \rangle - \|y\|_2 \\ &= \sup_{t \in \mathbb{R}_+} \sup_{y \in \mathbb{R}^n, \|y\|_2 = t} \langle y, x \rangle - \|y\|_2 \\ &= \sup_{t \in \mathbb{R}_+} t\|x\|_2 - t \\ &= \begin{cases} 0 & \text{si } \|x\|_2 \leq 1 \\ +\infty & \text{sinon} \end{cases} \end{aligned}$$
Le passage de la ligne 2 à la ligne 3 est possible en remarquant que le sup est atteint pour $y = t\frac{x}{\|x\|_2}$ (Cauchy-Schwarz).
5. Calculez f_3^*.
Solution : On a
$$f_3^*(x) = \sup_{y \in \mathbb{R}^n, \|y\|_2 \leq 1} \langle x, y \rangle - \frac{1}{2}\|y\|_2^2.$$
Si la contrainte est inactive dans ce problème, alors les conditions d'optimalité sont $y = x$ et $\|y\|_2 < 1$. Donc si $\|x\|_2 < 1$, $f_3^*(x) = \frac{1}{2}\|x\|_2^2$.
Si la contrainte est active, c'est-à-dire $\|x\|_2 \geq 1$, le problème devient :
$$f_3^*(x) = \sup_{y \in \mathbb{R}^n, \|y\|_2 = 1} \langle x, y \rangle - \frac{1}{2}\|y\|_2^2.$$
Dans ce cas, on réutilise le théorème de Cauchy-Schwarz pour conclure que l'optimum est $y = \frac{x}{\|x\|_2}$. Donc $f_3^*(x) = \|x\|_2 - \frac{1}{2}$.
Finalement, on voit que :
$$f_3^*(x) = \begin{cases} \frac{1}{2}\|x\|_2^2 & \text{si } \|x\|_2 \leq 1 \\ \|x\|_2 - \frac{1}{2} & \text{sinon.} \end{cases}$$

PARTIE 2 - Propriétés élémentaires de f^* :

1. Montrez que f^* est convexe.

Solution : On a $\forall \alpha \in [0,1]$,

$$\begin{aligned} f^*(\alpha x + (1-\alpha)x') &= \sup_{y \in \mathbb{R}^n} \langle \alpha x + (1-\alpha)x', y\rangle - f(y) && (4)\\ &= \sup_{y \in \mathbb{R}^n} \alpha\left(\langle x, y\rangle - f(y)\right) + (1-\alpha)\left(\langle x', y\rangle - f(y)\right) && (5)\\ &\leq \alpha \sup_{y \in \mathbb{R}^n} \left(\langle x, y\rangle - f(y)\right) + (1-\alpha) \sup_{y \in \mathbb{R}^n} \left(\langle x', y\rangle - f(y)\right) && (6)\\ &= \alpha f^*(x) + (1-\alpha) f^*(x'). && (7) \end{aligned}$$

2. On pose $h(x) = f(\lambda x)$, $\lambda \in \mathbb{R}^*$. Calculez $h^*(x)$ en fonction de f et f^*.
   Solution : On a

$$\begin{aligned} h^*(x) &= \sup_{y \in \mathbb{R}^n} \langle x, y\rangle - f(\lambda y) && (8)\\ &= \sup_{y' \in \mathbb{R}^n} \langle x, \frac{y'}{\lambda}\rangle - f(y') && (9)\\ &= f^*\left(\frac{x}{\lambda}\right). && (10) \end{aligned}$$

3. Que vaut $h^*(x)$ si $\lambda = 0$?
   Solution : Si $\lambda = 0$, $h(x) = f(0)$. Dans ce cas :

$$\begin{aligned} h^*(x) &= \sup_{y \in \mathbb{R}^n} \langle x, y\rangle - f(0) && (11)\\ &= \begin{cases} -f(0) & \text{si } x = 0 \\ +\infty & \text{sinon.} \end{cases} && (12) \end{aligned}$$

4. On souhaite montrer que $f^{**} = f$ lorsque f est convexe fermée.
   (a) Montrez que $f^{**}(x) = \sup_{y \in \mathbb{R}^n} \inf_{z \in \mathbb{R}^n} f(z) + \langle x - z, y\rangle$.

   Solution :

$$\begin{aligned} f^{**}(x) &= \sup_{y \in \mathbb{R}^n} \langle x, y\rangle - f^*(y) && (13)\\ &= \sup_{y \in \mathbb{R}^n} \langle x, y\rangle - \left(\sup_{z \in \mathbb{R}^n} \langle y, z\rangle - f(z)\right) && (14)\\ &= \sup_{y \in \mathbb{R}^n} \langle x, y\rangle + \left(\inf_{z \in \mathbb{R}^n} -\langle y, z\rangle + f(z)\right) && (15)\\ &= \sup_{y \in \mathbb{R}^n} \inf_{z \in \mathbb{R}^n} \langle x - z, y\rangle + f(z) && (16) \end{aligned}$$

   (b) En identifiant $(y, z) \mapsto f(z) + \langle x - z, y\rangle$ à un lagrangien, justifiez que l'on puisse intervertir le supremum et l'infimum.
   Solution : La fonction $\mathcal{L}(y, z) = f(z) + \langle x - z, y\rangle$ est convexe en z et concave en y. On peut donc intervertir le supremum et l'infimum.

   (c) Conclure que $f^{**} = f$.

Solution : On a donc :

$$\begin{aligned}
f^{**}(x) &= \sup_{y\in\mathbb{R}^n} \inf_{z\in\mathbb{R}^n} \langle x-z, y\rangle + f(z) && (17)\\
&= \inf_{z\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^n} \langle x-z, y\rangle + f(z) && (18)\\
&= \inf_{z\in\mathbb{R}^n} \begin{cases} f(z) & \text{si } \; x-z=0 \\ +\infty & \text{sinon} \end{cases} && (19)\\
&= f(z) && (20)
\end{aligned}$$

5. **Hors-Barème :** On considère le problème suivant :

$$(\mathcal{P}) \qquad \inf_{x\in\mathbb{R}^n} f(Ax) + g(x)$$

où $A \in \mathbb{R}^{m\times n}$ et f et g sont desz fonctions convexes fermées. En réécrivant le problème $(\mathcal{P})$ sous la forme

$$\inf_{x\in\mathbb{R}^n, y\in\mathbb{R}^m, Ax=y} f(y) + g(x)$$

et en utilisant la dualité lagrangienne, montrez que [1] :

$$\inf_{x\in\mathbb{R}^n} f(Ax) + g(x) = \sup_{\lambda\in\mathbb{R}^m} -g^*(A^T\lambda) - f^*(-\lambda).$$

Solution : Faites-vous les dents (en reprenant les idées ci-dessus) !

1. c'est un résultat central de la dualité de Fenchel-Rockafellar