Examen du cours d'optimisation différentiable

Durée : 2 heures
Les documents ainsi que les calculatrices ne sont pas autorisés.

****

Exercice 0.1 (transformée de Fenchel)
Soit $\phi : \mathbb{R}^d \to \mathbb{R}$. La transformée de Fenchel de ϕ est définie pour tout $u \in \mathbb{R}^d$ par

$$\phi^*(u) = \sup_{v\in\mathbb{R}^d} \left(\langle u, v\rangle - \phi(v)\right).$$

1. **Calculer la transformée de Fenchel de $\phi(v) = (1/2)\left\|v\right\|_2^2$ où $\left\|v\right\|_2$ est la norme ℓ_2^d de $v \in \mathbb{R}^d$.**
2. **On considère la fonction suivante : pour tout $v = (v_j)_{j=1}^d \in \mathbb{R}^d$,**

$$\phi(v) = \sum_{j=1}^{d} \exp(v_j).$$

2.1 Calculer la transformée ϕ^* de ϕ en tout point $u = (u_j)_{j=1}^d \in \mathbb{R}^d$ où $u_j > 0$ pour $j = 1, \ldots, d$.

2.2 Calculer $\nabla\phi^*(u)$ en tout point $u = (u_j)_{j=1}^d \in \mathbb{R}^d$ où $u_j > 0$ pour $j = 1, \ldots, d$.

2.3 Soit $u = (u_j)_{j=1}^d \in \mathbb{R}^d$ où $u_j > 0$ pour $j = 1, \ldots, d$. Calculer $\nabla\phi^*(\nabla\phi(u))$.

****

Correction de l'exercice 0.1 **1.** Soit $u \in \mathbb{R}^d$. On pose $f(v) = \langle u, v\rangle - \phi(v) = \langle u, v\rangle - (1/2)\left\|v\right\|_2^2$. On a $\nabla f(v) = u - v$ et $\nabla^2 f(v) = -I_d \prec 0$. Donc f est strictement concave. On a donc $v^* \in \mathrm{argmax}_{v\in} f(v)$ si et seulement si $\nabla(v^*) = 0$ càd $v^* = u$. On obtient alors

$$\phi^*(u) = \langle u, v^*\rangle - \phi(v^*) = \langle u, u\rangle - \frac{\left\|u\right\|_2^2}{2} = \frac{\left\|u\right\|_2^2}{2} = \phi(u).$$

2.1 Soit $u \in \mathbb{R}^d$ un vecteur à coordonnées strictement positives. On a

$$\phi^*(u) = \sup_{v\in\mathbb{R}^d} \left(\langle u, v\rangle - \sum_{j=1}^{d} \exp(v_j)\right) = \sup_{v\in\mathbb{R}^d} \left(\sum_{j=1}^{d} u_j v_j - \exp(v_j)\right) = \sum_{j=1}^{d} \sup_{v_j\in\mathbb{R}} \left(u_j v_j - \exp(v_j)\right).$$

Par ailleurs, comme $g : x \in \mathbb{R} \to tx - \exp(x)$ atteint son maximum en $x = \log(t)$ dès que $t > 0$ et vaut en ce point $t(\log(t) - 1)$, on a

$$\phi^*(u) = \sum_{j=1}^{d} u_j(\log(u_j) - 1).$$

2.2 On a pour tout $u \in \mathbb{R}^d$ à coordonnées strictement positives,

$$\nabla \phi^*(u) = (\log(u_j))_{j=1}^d.$$

2.3 On a alors pour tout $u \in \mathbb{R}^d$ à coordonnées strictement positives,

$$\nabla \phi^*(\nabla \phi(u)) = \nabla \phi^* \left((\exp(u_j))_{j=1}^d \right) = (u_j)_{j=1}^d = u.$$

****

Exercice 0.2

Soit le problème d'optimisation suivant

$$\min \left(yz + y^2 + z^2 : z \leq -1 \right). \tag{1}$$

1. **Réduire le problème de minimisation (1) à un problème de la forme**

$$\min_{t \in K} f(t)$$

où f est une fonction à valeurs réelles et

$$K = \{t : g(t) \leq 0\}$$

avec $g : \mathbb{R}^2 \to \mathbb{R}$ convexe de classe $\mathcal{C}^1$.

2. **Montrer que f est convexe.**
3. **Montrer que la contrainte K est qualifiée.**
4. **Écrire le lagrangien de (P).**
5. **Écrire le problème dual de (P) et le résoudre.**
6. **Écrire les conditions KKT de (P) et résoudre (P) à partir de ces équations.**

****

Correction de l'exercice 0.2

1. On considère le problème suivant :

$$\min_{t \in K} f(t)$$

où f est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ donnée par $f(y,z) = yz + y^2 + z^2$ et

$$K = \{(y,z)^\top \in \mathbb{R}^2 : g(y,z) \leq 0\}$$

avec $g(y,z) = z + 1$ convexe de classe $\mathbb{C}^1$

2. On a $f(y,z) = (y + z/2)^2 + 3z^2/4$. C'est une somme de carré de formes linéaires donc c'est une fonction convexe.
3. Il n'a y qu'une contrainte d'inégalité affine et la contrainte K est non vide donc la contrainte est qualifiée.

4. Le Lagrangien de (P) est donné pour tout $(y,z)^\top \in \mathbb{R}^2$ et $\lambda \geq 0$ par

$$L((y,z),\lambda) = f(y,z) + \lambda g(y,z) = yz + y^2 + z^2 + \lambda(z+1).$$

5. Le problème dual est

$$\sup_{\lambda \geq 0} d(\lambda) \text{ où } d(\lambda) = \min_{t \in \mathbb{R}^2} L(t,\lambda) = \frac{-\lambda^2}{3} + \lambda$$

   qui a pour solution $\lambda^* = 3/2$.

6. On rappelle que $(y^*, z^*, \lambda^*)^\top \in \mathbb{R}^3$ vérifient les conditions KKT du problème (2) quand
   (a) $(y^*, z^*)^\top \in K$, $\lambda^* \geq 0$,
   (b) $\nabla_{(y,z)} L((y^*, z^*), \lambda^*) = 0$,
   (c) $\lambda^* g(y^*, z^*) = 0$.

Le problème (2) est un problème d'OCD donc (y^*, z^*) est solution du problème (2) si et seulement si il existe λ^* une solution du problème dual tel que $(y^*, z^*, \lambda^*)^\top \in \mathbb{R}^3$ vérifient les conditions KKT. En résolvant les conditions KKT, on obtient que $y^* = 1/2$ et $z^* = -1$. Donc la valeur du problème d'optimisation est 1.75 atteint en $(y^*, z^*) = (1/2, -1)^\top$.

****

Exercice 0.3

Résoudre

$$\min\left(x_1^2 + x_2^2 + x_3^2 + x_4^2 : x_1 + x_2 + x_3 + x_4 = 1\right) \tag{2}$$

****

Correction de l'exercice 0.3 La fonction de Lagrange associée à (2) est définie en tout point $x = (x_j)_{j=1}^4 \in \mathbb{R}^4$ et $\lambda \in \mathbb{R}$ par

$$L(x,\lambda) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \lambda(x_1 + x_2 + x_3 + x_4 - 1).$$

Le problème dual associé est

$$\max_{\lambda \in \mathbb{R}} \Psi(\lambda) \text{ où } \Psi(\lambda) = \inf_{x \in \mathbb{R}^4} L(x,\lambda).$$

Comme $x \in \mathbb{R}^4 \to L(x,\lambda)$ est convexe et deux fois différentiable, $L(\cdot,\lambda)$ atteint son minimum sur $\mathbb{R}^4$ en un point x^* si et seulement si $\nabla L(x^*,\lambda) = 0$ càd pour $2x_i^* + \lambda = 0$ pour $i = 1,2,3,4$. On a alors

$$\Psi(\lambda) = L((-\lambda/2)_{i=1}^4, \lambda) = -\lambda^2 - \lambda.$$

Ainsi Ψ est maximale en $\lambda = -1/2$.

Le problème (2) est un problème d'OCD dont la contrainte est qualifiée vue que toutes les contraintes sont affines et la contrainte est non vide. On a donc, d'après KKT, l'équivalence entre les deux points suivant :

1. x^* est solution de (2)
2. il existe λ^* solution du problème dual tel que

i) $x^* \in K$

ii) $\nabla_x L(x^*, \lambda^*) = 0$

(Il n'y a pas de "complementary slackness condition" pour ce problème vu qu'il n'y a pas de contraintes d'inégalité).

On en déduit donc que x^* est solution de (2) si et seulement si

1. $x_1 + x_2 + x_3 + x_4 = 1$
2. $2x_i^* + \lambda^* = 0$ pour $i = 1, 2, 3, 4$ et $\lambda^* = -1/2$.

Ainsi l'unique solution de (2) est $(1/4)_{i=1}^4$.

Rem. : On aurait pu ne pas utiliser le fait que λ^* est solution du problème dual ici car la valeur de λ^* aurait pu être déduite de l'équation "$x_1 + x_2 + x_3 + x_4 = 1$".

****

Exercice 0.4 (Descente de gradient et régularisation)

Soit T fonctions <u>linéaires</u> $f_1, \ldots, f_T : \mathbb{R}^d \to \mathbb{R}$ et $\eta > 0$. On considère l'algorithme de descente de gradient suivant :

$$a_{t+1} = a_t - \eta \nabla f_t(a_t), \quad \forall t = 0, \ldots, T, \quad a_0 = 0.$$

Montrer que

$$a_{T+1} \in \operatorname*{argmin}_{a \in \mathbb{R}^d} \left(\sum_{t=1}^T f_t(a) + \frac{1}{2\eta} \|a\|_2^2 \right)$$

où on rappelle que $\|a\|_2$ est la norme ℓ_2^d de $a \in \mathbb{R}^d$ définie par

$$\|a\|_2 = \left(\sum_{j=1}^d a_j^2 \right)^{1/2}$$

et pour toute fonctions $f : \mathbb{R}^d \to \mathbb{R}$, $\operatorname{argmin}_{a \in \mathbb{R}^d} f(a) = \{a \in \mathbb{R}^d : f(a) \le f(b), \forall b \in \mathbb{R}^d\}$.

****

Correction de l'exercice 0.4 On note

$$f(a) = \sum_{t=1}^T f_t(a) + \frac{1}{2\eta} \|a\|_2^2 .$$

f est une fonction convexe de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^d$. Les minima de f sont les points critiques de f, càd

$$\operatorname*{argmin}_{a \in \mathbb{R}^d} f(a) = \left\{ a \in \mathbb{R}^d : \nabla f(a) = 0 \right\} .$$

Par ailleurs, on a

$$\nabla f(a) = \sum_{t=1}^T \nabla f_t(a) + \frac{a}{\eta}.$$

On a donc $a \in \operatorname{argmin}_{a \in \mathbb{R}^d} f(a)$ si et seulement si

$$a = -\eta \sum_{t=1}^T \nabla f_t(a).$$

De plus, les fonctions $f_t, t = 1, \dots, T$ sont linéaires, on a $f_t(a) = \langle g_t, a \rangle$ pour un certain g_t, on obtient que $a \in \operatorname{argmin}_{a \in \mathbb{R}^d} f(a)$ si et seulement si

$$a = -\eta \sum_{t=1}^{T} \nabla f_t(a) = -\eta \sum_{t=1}^{T} g_t.$$

Par ailleurs, l'algorithme de descente de gradient pour des fonctions linéaires est tel que

$$a_{T+1} = a_T - \eta \nabla f_T(a_T) = a_T - \eta g_T = -\eta \sum_{t=1}^{T} g_t$$

car $a_0 = 0$.