Rattrapage cours d'optimisation différentiable

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

****

Exercice 0.1 (Descente de gradient)
On considère la fonction

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

On remarque que $\varphi(x^*) = 0$ **si et seulement si** $x^* = 0$**. Décrire la convergence de l'algorithme de Newton pour résoudre l'équation** $\varphi(x) = 0$ **en fonction de la valeur du point initial.**

****

Correction de l'exercice 0.1 On a pour tout $t \in \mathbb{R}$,

$$\varphi'(t) = (1-t^2)^{-3/2}.$$

Alors l'algorithme de Newton est $(t_k)_k$ pour une certain point initial $t_0 \in \mathbb{R}$ et

$$t_{k+1} = t_k - \frac{\varphi(t_k)}{\varphi'(t_k)} = -t_k^3.$$

On a donc

1. $(t_k)_k$ converge vers 0 quand $|t_0| < 1$
2. $(t_k)_k$ est constante en 1 ou oscille entre -1 et 1 quand $|t_0| = 1$
3. $(t_k)_k$ diverge quand $|t_0| > 1$.

****

Exercice 0.2 (Système d'équations linéaire et optimisation)
Soit A **une matrice symétrique définie positive de** $\mathbb{R}^{N\times N}$ **et** $y \in \mathbb{R}^N$**.**

1. **Montrer que les deux assertions suivantes sont équivalentes :**
   (a) $\hat{x}$ **est solution de** $Ax = y$
   (b) $\hat{x}$ **minimise la fonction** $x \to F(x) = \frac{1}{2}x^\top Ax - x^\top y$**.**
2. **En utilisant *1.*, proposer un algorithme itératif pour approcher une solution du système** $Ax = y$ **ne faisant pas intervenir l'inverse de** A **ou la résolution d'un système d'équations linéaires.**

****

Correction de l'exercice 0.2 **1.** Première solution : La Hessienne de F est A qui est symétrique définie positive. Donc $\hat{x}$ minimise F si et seulement si $\nabla F(\hat{x}) = 0$. Or $\nabla F(x) = Ax - y$.

Deuxième solution : Si $Ax_0 = y$ alors pour tout $x \in \mathbb{R}^N$, on a

$$F(x) - F(x_0) = \tfrac{1}{2}x^\top Ax - x^\top y - \tfrac{1}{2}x_0^\top Ax_0 + x_0^\top y = \tfrac{1}{2}\Big(x^\top Ax + x_0^\top Ax_0 - 2x^\top Ax_0\Big). \tag{1}$$

Or par Cauchy-Schwartz, on a

$$2|x^\top Ax_0| = 2|\big\langle A^{1/2}x, A^{1/2}x_0\big\rangle| \leq 2\left\|A^{1/2}x\right\|_2 \left\|A^{1/2}x_0\right\|_2 \leq \left\|A^{1/2}x\right\|_2^2 + \left\|A^{1/2}x_0\right\|_2^2 = x^\top Ax + x_0^\top Ax_0.$$

On en déduit que $F(x) - F(x_0) \geq 0$. Donc x_0 est bien le minimum de F.

Réciproquement, si pour tout x, $F(x) - F(x_0) \geq 0$ alors pour tout $h \in \mathbb{R}^N$, on a

$$F(x_0 + h) = F(x_0) + \big\langle \nabla F(x_0), h\big\rangle + \frac{1}{2}h^\top \nabla^2 F(x_0)h = F(x_0) + \big\langle Ax_0 - y, h\big\rangle + \frac{\left\|A^{1/2}h\right\|_2^2}{2}$$

et comme $F(x_0+h) \geq F(x_0)$, on en déduit que pour tout h, $2\big\langle Ax_0 - y, h\big\rangle + \left\|A^{1/2}h\right\|_2^2 \geq 0$. Et donc, quitte à remplacer h par λh pour $\lambda > 0$, on voit que $\big\langle Ax_0 - y, h\big\rangle \geq 0$ pour tout h donc $Ax_0 = y$.

2. D'après *1.*, pour résoudre $Ax = y$ il suffit de trouver un minimum de la fonction $x \to F(x) = \frac{1}{2}x^\top Ax - x^\top y$. Pour cela on peut considérer un algorithme de Newton ou un algorithme de descente de gradient.

Newton s'écrit ici

$$x_{k+1} = x_k - (\nabla^2 F(x_k))^{-1}(\nabla F(x_k)) = x_k - A^{-1}(Ax_k - y)$$

et nécessite donc l'inversion de A. Un algorithme de descente de gradient s'écrit ici

$$x_{k+1} = x_k - \eta_k \nabla F(x_k) = x_k - \eta_k(Ax_k - y)$$

où $(\eta_k)_k$ est une suite de réels. La descente de gradient ne nécessite donc pas d'inverser A ou de résoudre une système d'équations linéaires à chaque étape.

****

Exercice 0.3
On définit

$$\Lambda = \Big\{\theta = (\theta_j)_{j=1}^d \in \mathbb{R}^d : \theta_j \geq 0, \forall j = 1, \ldots, d\Big\}.$$

Soit $x \in \mathbb{R}^d$. Calculer la valeur

$$(\star) = \inf_{\theta \in \Lambda} \frac{1}{2}\sum_{j=1}^d \left(\theta_j x_j^2 + \frac{1}{\theta_j}\right).$$

****

Correction de l'exercice 0.3 On va montrer que

$$\|x\|_1 = \inf_{\theta\in\Lambda} \frac{1}{2}\sum_{j=1}^d \left(\theta_j x_j^2 + \frac{1}{\theta_j}\right).$$

où $\|x\|_1 = \sum_{j=1}^d |x_j|$ est la norme ℓ_1^d de x.

On a

$$\inf_{\theta\in\Lambda} \frac{1}{2}\sum_{j=1}^d \left(\theta_j x_j^2 + \frac{1}{\theta_j}\right) = \frac{1}{2}\sum_{j=1}^d \inf_{\theta_j\geq 0}\left(\theta_j x_j^2 + \frac{1}{\theta_j}\right).$$

De plus, pour tout $t\in\mathbb{R}$, $\theta\geq 0 \to \theta t^2 + 1/\theta$ est minimal en $\theta = 1/|t|$. Donc,

$$\inf_{\theta\in\Lambda} \frac{1}{2}\sum_{j=1}^d \left(\theta_j x_j^2 + \frac{1}{\theta_j}\right) = \frac{1}{2}\sum_{j=1}^d \inf_{\theta_j\geq 0}\left(\theta_j x_j^2 + \frac{1}{\theta_j}\right) = \frac{1}{2}\sum_{j=1}^d \left(\frac{x_j^2}{|x_j|} + \frac{1}{1/|x_j|}\right) = \sum_{j=1}^d |x_j| = \|x\|_1.$$

****

Exercice 0.4 (KKT en optimisation différentiable)

Trouver une solution au problème

$$\max\left(xy : x + y^2 \leq 2, x\geq 0, y\geq 0\right) \tag{2}$$

****

Correction de l'exercice 0.4 On pose

$$f : \begin{cases} \mathbb{R}^2 & \to & \mathbb{R} \\ (x,y) & \to & -xy \end{cases}$$

et pour tout $(x,y)\in\mathbb{R}^2$, $h_1(x,y) = x + y^2 - 2$, $h_2(x,y) = -x$ et $h_3(x,y) = -y$. On introduit aussi la contrainte K définie par

$$K = \left\{(x,y)\in\mathbb{R}^2 : h_i(x,y)\leq 0 \text{ pour } i = 1,2,3\right\}.$$

Le problème (2) est équivalent à

$$\min_{x\in K} f(x). \tag{3}$$

Comme la fonction-objectif f est continue et que la contrainte K est compacte, d'après Weierstrass, le problème (3) admet au moins une solution.

La contrainte K est qualifiée par la condition de Slater : il n'y a pas de contraintes d'égalité, les contraintes d'inégalités sont convexes et $\mathcal{C}^1$ et il existe $(x_0,y_0) = (1/2,1/2)\in K$ tel que $h_i(x_0,y_0) < 0$ pour $i = 1,2,3$. On peut donc appliquer le théorème de KKT sur toute la contrainte K.

Par ailleurs, f et h_1,h_2,h_3 sont de classe $\mathcal{C}^1$ donc d'après KKT, si (x^*,y^*) est solution de (3) alors il existe $\lambda_1,\lambda_2,\lambda_3$ tels que

1. $(x^*,y^*)\in K$ et $\lambda_j\geq 0$, pour $j = 1,2,3$
2. $\nabla f(x^*,y^*) + \lambda_1\nabla h_1(x^*,y^*) + \lambda_2\nabla h_2(x^*,y^*) + \lambda_3\nabla h_3(x^*,y^*) = 0$

3. $\lambda_j h_j(x^*, y^*) = 0$ pour tout $j = 1, 2, 3$.

   On a pour tout $(x, y) \in \mathbb{R}^2$,

$$\nabla f(x,y) + \lambda_1 \nabla h_1(x,y) + \lambda_2 \nabla h_2(x,y) + \lambda_3 \nabla h_3(x,y) = \begin{pmatrix} -y + \lambda_1 - \lambda_2 \\ -x + 2\lambda_1 y - \lambda_3 \end{pmatrix}$$

Alors le point *2.* de KKT est équivalent à

$$\begin{cases} -y + \lambda_1 - \lambda_2 = 0 \\ -x + 2\lambda_1 y - \lambda_3 = 0 \end{cases} \text{càd} \begin{cases} y = \lambda_1 - \lambda_2 \\ x = 2\lambda_1(\lambda_1 - \lambda_2) - \lambda_3 \end{cases}$$

D'après les "complementary condition", on a soit $x = 0$ ou $\lambda_2 = 0$ et dans le cas où $x = 0$, $f(x, y) = 0$, soit $y = 0$ ou $\lambda_3 = 0$ et dans le cas où $y = 0$ on a aussi $f(x, y) = 0$. On suppose alors que $x > 0$ et $y > 0$ et on regarde si la valeur de la fonction objectif pour les solutions obtenues dans ce cas sont plus petites que 0. On a donc $\lambda_2 = \lambda_3 = 0$ alors une solution de (3) vérifie

$$\begin{cases} y = \lambda_1 \\ x = 2\lambda_1^2 \end{cases}$$

càd $x = 2y^2$. Par ailleurs, d'après la "complementary condition", on a soit $\lambda_1 = 0$ soit $x + y^2 = 2$. Dans le premier cas, on a $x = y = 0$ et dans le deuxième cas, on a $x = 2y^2$ et $2 = x + y^2 = 3y^2$ donc $y = \sqrt{2/3}$ et $x = 4/3$.

On en déduit que f atteint son minimum sur K soit en un point de la forme $(0, y)$ ou $(x, 0)$ ou en $(4/3, \sqrt{2/3})$. Dans les deux premiers cas, la fonction objectif vaut 0 et dans l'autre cas elle vaut $-(4/3)\sqrt{2/3}$. Il y a donc une unique solution au problème (2) qui est $(4/3, \sqrt{2/3})$.