Université Paris Descartes
UFR de Mathématiques et Informatique
45, rue des Saints-Pères 75270 Paris cedex 06

UNIVERSITÉ PARIS DESCARTES

MASTER I 2012–2013

OPTIMISATION ET ALGORITHMIQUE

EXAMEN, durée 2h

Tout document est interdit.
Tout appareil électronique, même à titre d'horloge, est également interdit.
Nombre de pages de l'énoncé : 2
Il faut justifier les réponses. Il sera tenu compte de la rédaction de la copie.

Exercice 1

On considère la fonction f, définie pour tout $x \in \mathbb{R}^n$ par $f(x) = \dfrac{1}{2}\displaystyle\sum_{k=1}^{m} r_k(x)^2$

où $r_k \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ pour $1 \leq k \leq m$. On note

$$r(x) = \begin{pmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{pmatrix} \in \mathbb{R}^m, \qquad J(x) = \begin{pmatrix} \frac{\partial r_1}{\partial x_1}(x) & \cdots & \frac{\partial r_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial r_m}{\partial x_1}(x) & \cdots & \frac{\partial r_m}{\partial x_n}(x) \end{pmatrix} \in \mathcal{M}(m,n)$$

et $\|.\|_2$ la norme euclidienne associée au produit scalaire canonique $< .,. >$.

On suppose $m \geq n$ et que pour tout $x \in \mathbb{R}^n$, $\text{rang}(J(x)) = n$.
Pour minimiser f, on propose une méthode de descente inspirée de la méthode de Newton.

1. Calculer le gradient $\nabla f(x)$ et l'exprimer grâce à $J(x)^t$ et $r(x)$.
   Calculer la hessienne $H_f(x)$ et montrer que l'on peut écrire $H_f(x) = A(x) + S(x)$ où $A(x) = J(x)^t\, J(x)$ et $S(x)$ est une somme (à déterminer) composée des $r_k(x)$ et $H_{r_k}(x)$, $1 \leq k \leq m$.

2. Pour $x \in \mathbb{R}^n$, on note $\sigma_1(x) \geq \cdots \geq \sigma_n(x)$ les valeurs singulières de $J(x)$.
   Pourquoi y en a-t-il n ? que peut-on dire de $\sigma_n(x)$?

3. On détermine la direction d, en x, comme solution du système $A(x)\, d = -\nabla f(x)$.
   Montrer que la matrice $A(x) = J(x)^t J(x)$ est symétrique définie positive.
   Montrer que d est une direction de descente de f en x si $\nabla f(x) \neq 0$.
   Comparer à la méthode de NEWTON .

4. Écrire l'algorithme de minimisation qui utilise la direction de descente du 3) et qui part d'un point $x^{(0)} \in \mathbb{R}^n$.
   Note : on suppose avoir une méthode pour déterminer un pas σ_k vérifiant les conditions de Wolfe.

5. Soit $x^{(0)} \in \mathbb{R}^n$ la valeur initiale de l'algorithme et $L_f(x^{(0)}) = \{x \in \mathbb{R}^n \,/\, f(x) \leq f(x^{(0)})\}$.
   On suppose qu'il existe $\gamma \in \mathbb{R}_+^*$ tel que $\inf\limits_{x \in L_f(x)} \sigma_n(x) \geq \gamma$,
   où $\sigma_n(x)$ est la plus petite valeur singulière de $J(x)$.
   Soit $x \in L_f(x^{(0)})$ et d la direction de descente en x, on pose $\cos\theta = -\dfrac{< \nabla f(x), d >}{\|\nabla f(x)\|_2 \; \|d\|_2}$.
   Monter qu'il existe $\delta > 0$, indépendant de x , tel que $\cos\theta > \delta$.

6. Interprétez le résultat précédent en termes de convergence de la méthode de descente proposée.

Exercice 2

Soit $A = (a_{ij})_{1\le i,j\le d}$ une matrice réelle donnée. On veut résoudre le problème d'optimisation suivant

$$x^\star = \arg\min_{x\in\mathbb{R}^d} f(x)\,, \qquad \text{où } \; f(x) = \ln\left(\sum_{1\le i,j\le d} a_{ij}^2\, e^{x_i - x_j}\right)\,.$$

On note $\{e_1,\dots,e_d\}$ la base canonique de $\mathbb{R}^d$; $<.,.>$ le produit scalaire canonique sur $\mathbb{R}^d$ et pour $x = (x_1,\dots,x_d)^t \in \mathbb{R}^d$, $\|x\|_1 = \sum_{i=1}^{d} |x_i|$ et $\|x\|_\infty = \max_{1\le i\le d} |x_i|$.

1. Soit $g \in \mathbb{R}^d$ fixé. Montrer que pour tout $v \in \mathbb{R}^d$, $\|v\|_1 \le 1$: $|<g,v>| \le \|g\|_\infty$.
   Pour quels vecteurs v, $\|v\|_1 \le 1$, a-t-on égalité ?
   En déduire que $\arg\max_{v\in\mathbb{R}^d, \|v\|_1\le 1} |<g,v>| = \pm e_k$, avec $k \in \{1,\dots,d\}$ que l'on déterminera.

2. Pour $k \in \{1,\dots,d\}$ fixé, montrer que l'on peut écrire f sous la forme

   $$f(x) = \ln(\alpha_k + \beta_k e^{-x_k} + \gamma_k e^{x_k})$$

   où α_k, β_k et γ_k sont des quantités positives indépendantes de x_k.

3. Pour $x \in \mathbb{R}^d$ donné et $k \in \{1,\dots,d\}$, on pose, pour $\sigma \in \mathbb{R}$: $\phi_k(\sigma) = f(x + \sigma\, e_k)$.
   Calculer $\phi_k'(\sigma)$. Déterminer la solution $\sigma^\star$ de l'équation $\phi_k'(\sigma) = 0$.
   En fonction du signe de $\phi_k'(0)$ déterminez celui de $\sigma^\star$ et dressez le tableau de variations de ϕ_k dans les deux cas.

4. En se basant sur ce qui précède, on propose d'utiliser l'algorithme suivant pour minimiser f :

   choisir $x^{(0)}$, poser $p = 0$

   tant que $\left\|\nabla f(x^{(\mathrm{p})})\right\| > \varepsilon$

   choisir $k \in \{1,\dots,d\}$ tel que $\left|\dfrac{\partial f}{\partial x_k}(x^{(\mathrm{p})})\right| = \|\nabla f(x^{(\mathrm{p})})\|_\infty$

   calculer σ_p, solution de $\phi_k'(\sigma) = 0$, où $\phi_k(\sigma) = f(x^{(\mathrm{p})} + \sigma\, e_k)$

   poser $x^{(\mathrm{p}+1)} = x^{(\mathrm{p})} + \sigma_p\, e_k$

   faire $p = p + 1$

   (a) Comparez $x^{(\mathrm{p}+1)}$ et $x^{(\mathrm{p})}$. Que pouvez vous en dire ?

   (b) Grâce aux résultats précédents, donnez une justification et interprétation de l'algorithme en tant que méthode de descente (choix de la direction et du pas de descente, vérification du fait que l'on a une direction de descente, complexité,...).

   (c) Quels avantages peut-on trouver à cette méthode par rapport à la méthode de descente du gradient ? la méthode de Newton ?