Université de Franche-Comté
Département de Mathématiques Février 2009

Recherche Opérationnelle - M1
Examen de fin d'année

Exercice 1. On considère la fonction :

$$\begin{array}{rccc} f & : & \mathbb{R}^2 & \to & \mathbb{R} \\ & & (x,y) & \mapsto & \frac{1}{4}\left(e^x + y^2\right) \end{array}$$

ainsi que l'ensemble $C = \{(x,y) \in \mathbb{R}^2 : \log(2) \leq x \leq \log(20) \text{ et } 1 \leq y \leq 4\}$.

1. La fonction f est-elle convexe ?
2. Rappeler rapidement pourquoi l'ensemble C est convexe.
3. En s'inspirant du dessin, proposer un candidat $z^\star$ au problème de minimisation de f sur C.
4. Représenter sur le dessin le cône normal à C en $z^\star$ ainsi que l'opposé du gradient en $z^\star$.
5. Résoudre le problème mathématiquement.

Exercice 2 (Entropie et optima)**.** On s'intéresse à l'entropie et à ce que les physiciens appellent la mesure de Gibbs. C'est un concept assez simple mais fondamental pour plein d'applications statistiques nécéssitant une approche Bayesienne avec usage d'outils probabilistes comme ce qu'on appelle des Chaînes de Markov.

Supposons qu'on ait un ensemble E fini de cardinal n. Par exemple, E est l'ensemble des modèles pour un mélange de K régressions logistiques (K fixé pour simplifier), un modèle étant par exemple la donnée de qui sont les coefficients de régression qui sont nuls comme cela arrive souvent en pratique. Un autre exemple est l'ensemble des colorations permises d'un graphe donné, comme utilisé dans les télécoms pour l'affectation de fréquences. Je donne juste ces exemples pour donner une idée de ce à quoi cela peut servir mais on ne les utilisera pas dans la suite.

On peut mettre une loi uniforme sur E mais il y a beaucoup de cas, notamment en optimisation et en statistiques Bayesiennes, où on veut mettre autre chose que la loi uniforme. Pour choisir une loi raisonnable, les physiciens (Boltzmann par exemple) ont introduit la notion d'entropie associée à une loi de probabilité discrète sur E. Soit π une loi de probabilité sur E, l'entropie H de π est donnée par

$$H(\pi) = -\sum_{i=1}^{n} \pi_i \log(\pi_i) \tag{1}$$

avec la convention $0\log(0) = 0$. En physique, on est intéressé par les notions d'énergie des systèmes dans une situation donnée. Pour nous, une énergie est une fonction f positive sur E. L'énergie moyenne $\mu_f(\pi) = \sum_{i=1}^{n} f(i)\pi_i$ est une quantité informative.

Le but de cet exercice est de trouver le vecteur de probabilités π^* qui maximise l'entropie sous la contrainte que $\mu_f(\pi) = \mu$ avec μ un réel positif donné.

1. L'entropie est-elle une fonction convexe ?
2. Ecrire le problème de maximisation correspondant à la recherche du π optimal. (N.B. : on n'aura pas besoin d'inclure les contraintes $\pi_i \geq 0$ vu qu'elle sont déjà prises en compte par l'utilisation de la fonction logarithme.)
3. Ecrire la fonction de Lagrange. Quel est l'ensemble K ? quel est son polaire K° ?
4. La fonction de Lagrange est-elle concave, vue comme fonction de π pour $u \in K^\circ$?
5. Ecrire la fonction duale de Lagrange $\theta(u)$. Cette fonction est-elle convexe ?
6. Ecrire le problème dual. L'optimum de la fonction duale est-il unique ?
7. La fonction duale est-elle différentiable en l'optimum ?
8. En déduire la forme de la solution $\pi^\star$ en fonction de la solution duale $u^\star$.

La loi obtenue est appelée mesure de Gibbs associée à l'energie μ.

Exercice 3 (Correction des moindres carrés pour les observations astronomiques)**.** Beaucoup de données en astronomie sont des données spectrales : on regarde un amas stellaire par exemple et on mesure l'amplitude du signal émis pour une certaine gamme de fréquences : plus précisément pour les fréquences $f_1,\dots,f_m$, on observe sur notre antenne les amplitudes $A(f_1),\dots,A(f_m)$.

Le problème des astronomes est de modéliser la relation entre A et f. Pour cela, ils utilisent des fonctions de base, comme des sinus/cosinus, des exponentielles, des splines, des ondelettes etc... et expriment A comme :

$$A(f) = \sum_{i=1}^{n} x_i \psi_i(f) + \varepsilon(f), \tag{2}$$

où les ψ_i sont les fonctions choisies et mises dans l'ordre qu'ils souhaitent et $\varepsilon(f)$ est une variable de f gaussienne et de variance σ^2 supposée indépendante de f. En notant par a le vecteur des observations $A(f_1),\dots,A(f_m)$ et Ψ la matrices des $\psi_i(f_j)$, on a donc :

$$a = \begin{bmatrix} A(f_1) \\ \vdots \\ A(f_m) \end{bmatrix} = \begin{bmatrix} \psi_1(f_1) & \dots & \psi_n(f_1) \\ \vdots & & \vdots \\ \psi_1(f_m) & \dots & \psi_n(f_m) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix} = \Psi x + \varepsilon, \tag{3}$$

où on suppose que ε a des composantes indépendantes. Le problème est de déterminer x de manière optimale. Bien entendu, une façon raisonnable et bien connue est la méthode des moindes carrés qui consiste à minimiser les carrés des écarts à l'observation, c'est-à-dire

$$\min_{x \in \mathbb{R}^n} \|a - \Psi x\|_2^2. \tag{4}$$

Un gros problème en astronomie est que les $A(f_j)$ tombent principalement dans deux catégories bien distinctes : les petites valeurs non-nulles correspondent aux étoiles et sont dans le domaine rouge du spectre et les grandes valeurs non-nulles correspondent à la partie "continue" de l'amas (gaz, poussières, etc...). En pratique, on observe que l'objectif des moindes carrés va tendre à avoir une approximation la meilleure possible pour la partie des grandes valeurs de $A(f_j)$ en négligeant un peu ce qui se passe pour les petites valeurs de $A(f_j)$. Or ce qui se passe pour les petites valeurs (aux fréquences dans le rouge) est très important car cela permet de connaître le nombre d'étoiles dans l'amas. Il faut donc corriger cet estimateur des moindres carrés.

Pour cela, on introduit la contrainte supplémentaire que l'erreur commise sur les fréquences $f_1, \ldots, f_p$ correspondant au rouge (donc aux étoiles) doit être inférieure à α, c'est-à-dire $\|a_{1:p} - \Psi_{1:p}x\|_2^2 \leq \alpha$ où $a_{1:p}$ est le vecteur des observations $A(f_1),\ldots,A(f_p)$ et $\Psi_{1:p}$ est la sous matrice de Ψ correspondant aux p premières lignes. On obtient ainsi un problème de régression sous contrainte :

$$\min_{x\in\mathbb{R}^n} \|a - \Psi x\|_2^2 \qquad \text{t.q.} \qquad \|a_{1:p} - \Psi_{1:p}x\|_2^2 \leq \alpha. \tag{5}$$

1. Le problème (5) est-il convexe ? Est-il un problème de programmation quadratique ?
2. Soit x^* une solution de ce problème. Ecrire la condition d'optimalité de Kuhn et Tucker en x^*. Les conditions de qualification des contraintes de Kuhn et Tucker sont-elles vérifiées dans notre problème ?
3. Donner la fonction de Lagrange $L(x, u)$ associée à ce problème. Quel est le cône K ici ? Quel est son polaire K° ?
4. La fonction de Lagrange est-elle concave (vue comme fonction de x) pour $u \in K^\circ$? Calculer le gradient de la fonction de Lagrange (vue comme fonction de x).
5. En déduire la fonction duale de Lagrange $\theta(u)$. On pose $q(u) = \Psi^t a + u\Psi_{1:p}^t a_{1:p}$ et $Q(u) = \Psi^t\Psi + u\Psi_{1:p}^t\Psi_{1:p}$. Vérifier que :

   $$\theta(u) = q(u)^t Q(u)^{-1} q(u) - a^t a - u a_{1:p}^t a_{1:p} + u\alpha.$$

   Cette fonction θ est-elle convexe ?
6. Quelle est la différentielle de $u \mapsto q(u)$ en u_0, un réel quelconque ? et celle de $u \mapsto q(u)^t$ en u_0 ? (On précisera l'ensemble de départ et d'arrivée pour $u \mapsto q(u)$ et $u \mapsto q^t(u)$ pour s'aider).
7. Quel est le problème dual ?
8. La fonction $X \mapsto X^{-1}$ pour les matrices inversibles est-elle différentiable ? (Pas besoin de calculer la differentielle explicitement ici : on pourra s'aider des Formules de Cramer mais répondre en quelques lignes sans faire de grands calculs surtout ...). Peut-on en déduire que la fonction duale θ est différentiable pour $u \in K^\circ$?
9. En particulier, la fonction duale θ est-elle différentiable à l'optimum ? Qu'en concluez vous ?

   N.B. : Le reste de cet exercice consiste à exprimer explicitement la différentielle de θ et risque d'être assez sportif si vous êtes allergique aux différentielles. Si c'est le cas, passez directement à la question 12. en supposant qu'on connaisse la dérivée de θ et retentez à la fin si vous avez un peu de temps !

10. Soient les fonction $f : X \mapsto X$ et $g : X \mapsto X^{-1}$ définies sur les matrices symétriques réelles définies positives. Comme on a $fg = I$ on obtient par la règle de différentiation du produit $df_{X_0}(H)g(X_0) + f(X_0)dg_{X_0}(H) = 0$. En déduire que la différentielle de g en X_0, une matrice quelconque, est $H \mapsto -X_0^{-1}HX_0^{-1}$.
11. Déduire des deux questions précédentes que la différentielle de θ en u_0 est donnée par :

    $$h \mapsto (2a_{1:p}^t\Psi_{1:p}Q(u_0)^{-1}q(u_0) - q(u_0)^t Q(u_0)^{-1}\Psi_{1:p}^t\Psi_{1:p}Q(u_0)^{-1}q(u_0) - a_{1:p}^t a_{1:p} + \alpha)\, h. \tag{6}$$

    Donner alors la dérivée de θ en u_0.

12. Il s'agit alors de trouver le u^* tel que la dérivée de θ en u^* est nulle. Est-ce faisable à la main ? Sinon, donner le principe de la méthode du gradient pour trouver l'optimum.
13. (Question bonus 1) Calculer la dérivée seconde de θ. Décrire succinctement alors la méthode de Newton pour trouver u^*.
14. (Question bonus 2) Si x est un vecteur très grand, la formulation duale de ce problème a un énorme avantage par rapport à toute méthode qui attaquerait le problème de départ frontalement : lequel ?

Exercice 4 (Contrainte sur les r plus grandes composantes)**.** Il arrive souvent qu'on veuille imposer que la somme des r plus grandes composantes d'un vecteur $x \in \mathbb{R}^n$ soit inférieure à un niveau donné α. le but de cet exercice est de comprendre comment formuler simplement ce type de contraintes.

Soit $x \in \mathbb{R}^n$ et $x_{(1)}, \ldots, x_{(n)}$ ses composantes réordonnées dans l'ordre décroissant. On a donc $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$. La contrainte qui nous intéresse se formule donc comme :

$$x_{(1)} + \cdots + x_{(r)} \leq \alpha. \tag{7}$$

1. Montrer que $x_{(1)} + \cdots + x_{(r)}$ est la valeur optimale du programme linéaire avec contraintes binaires suivant (x est fixé) :

$$\max_{y \in \mathbb{R}^n} x^t y \qquad \text{t.q. } \sum_{i=1}^{n} y_i = r \text{ et } y_i \in \{0, 1\},\ i = 1, \ldots, n. \tag{8}$$

2. Montrer que le programme linéaire suivant a une solution dont toutes les composantes sont égales à zéro ou un :

$$\max_{y \in \mathbb{R}^n} x^t y \qquad \text{t.q. } \sum_{i=1}^{n} y_i = r \text{ et } y_i \in [0, 1],\ i = 1, \ldots, n. \tag{9}$$

   La solution que vous avez considérée est-elle solution du problème linéaire avec contraintes binaires de la question précédente ?
3. Donner le problème dual au problème de la question précédente. Montrez qu'il peut se mettre sous la forme :

$$\min_{u,v} \sum_{i=1}^{n} u_i + rv \qquad \text{t.q. } v + u_i \geq x_i,\ i = 1, \ldots, n \text{ et } u \geq 0. \tag{10}$$

   (On pourra utiliser le vecteur e dont toutes les composantes sont égales à 1 pour simplifier les calculs.)
4. La dualité forte pour les programmes linéaires est un résultat puissant qui dit qu'un programme linéaire et son dual ont la même valeur optimale. Ce résultat permet de caractériser le fait que $x_{(1)} + \cdots + x_{(r)} \leq \alpha$. Montrer en effet que $x_{(1)} + \cdots + x_{(r)} \leq \alpha$ si et seulement si il existe $u \in \mathbb{R}^n$ et $v \in \mathbb{R}$ tels que $\sum_{i=1}^{n} u_i + rv \leq \alpha, \quad v + u_i \geq x_i,\ i = 1, \ldots, n$ et $u \geq 0$. Qu'aurait-on pu dire si on n'avait eu à notre disposition que la propriété de dualité faible ?
5. **Application :** Dans le problème du choix de portefeuille de Markowitz, on a souvent envie d'imposer la contrainte supplémentaire que 10% des produits choisis ne peuvent pas représenter plus de 80% de l'investissement. Exprimer le problème de Markowitz prenant en compte cette contrainte de diversification comme un programme quadratique en vous appuyant sur le résultat de la question précédente.