Ensimag 2A – automne 2014
Examen du vendredi 20 décembre
Alejandro Blumentals, Florent Cadoux, Jérôme Malick

# Optimisation numérique - Examen déc. 2014

– Tout document papier est autorisé.
– L'énoncé fait quatre pages. Les deux exercices sont indépendants.
– Barème approximatif : 4 – 16.
– La qualité de la présentation et de l'argumentation sera un élément important de la notation.

**Exercice 1 – Algorithmes, implémentation.** Donner des réponses (brèves) aux questions de cours/TPs suivantes :

**a)** **(Recherche linéaire)** Expliquer quel est le rôle d'une recherche linéaire. Quel était l'impact de la recherche linéaire de Wolfe sur la méthode de gradient appliquée aux fonctions du TP1-2-3 ? Pour une fonction différentiable, peut-on l'utiliser pour n'importe quelle direction ? Pour une fonction non-différentiable, peut-on l'utiliser avec un sous-gradient ?

**b)** **(Test d'arrêt)** Quel est le test d'arrêt naturel en optimisation différentiable ? Est-il valide aussi en optimisation non-différentiable ? En pratique, quel autre test ajoute-t-on lors de la programmation d'une méthode itérative ?

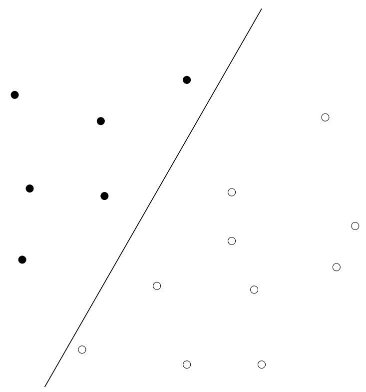
**c)** **(Expériences)** Quels types de tests peut-on faire pour valider un algorithme d'optimisation ? Pour évaluer son efficacité ? Et pour comparer différents algorithmes entre eux ?

**d)** **(Faisceaux)** Quel type de sous-problème donne l'itéré suivant dans la méthode des faisceaux ? Quelles sont les modifications à faire sur ce sous-problème d'une itération à la suivante ? Quel sont ses avantages par rapport au sous-problème de la méthode de Kelley ?

**Exercice 2 – Discrimination par SVM.** Cet exercice porte sur une méthode classique de l'apprentissage statistique (SVM pour "support vector machine").

Soient $m$ points $x_1, \dots, x_m$ de $\mathbb{R}^n$, distribués autour de 0. On suppose qu'ils se partagent en deux ensembles avec chacun une caractéristique commune – ce qui se représente de la manière suivante : on a deux ensembles d'indices $I_+$ et $I_-$, et à chaque point $x_i \in \mathbb{R}^n$ ($i = 1, \dots, m$) est associé un booléen $y_i \in \{-1, +1\}$ tel que $y_i = 1$ si $i \in I_+$ et $y_i = -1$ si $i \in I_-$.

On souhaite séparer ces deux ensembles de points par un hyperplan, ce qui revient à chercher un vecteur $w \in \mathbb{R}^n$ tel que $w^\top x_i > 0$, pour tout $i \in I_+$, et $w^\top x_i < 0$ pour tout $i \in I_-$. Géométriquement, l'hyperplan $H$ orthogonal à $w$ sépare ainsi l'espace $\mathbb{R}^n$ en deux avec $I_+$ et $I_-$ chacun d'un côté (comme illustré ci-contre sur le dessin dans $\mathbb{R}^2$, avec les points de $I_+$ en noir à gauche et ceux de $I_-$ en blanc à droite). On peut ainsi assigner un nouveau $\bar{x}$, sans caractéristique, à l'un des deux ensembles en fonction de sa position dans l'espace : si $w^\top \bar{x} > 0$, alors on considère que $\bar{x}$ est dans $I_+$, sinon on considère qu'il est dans $I_-$.

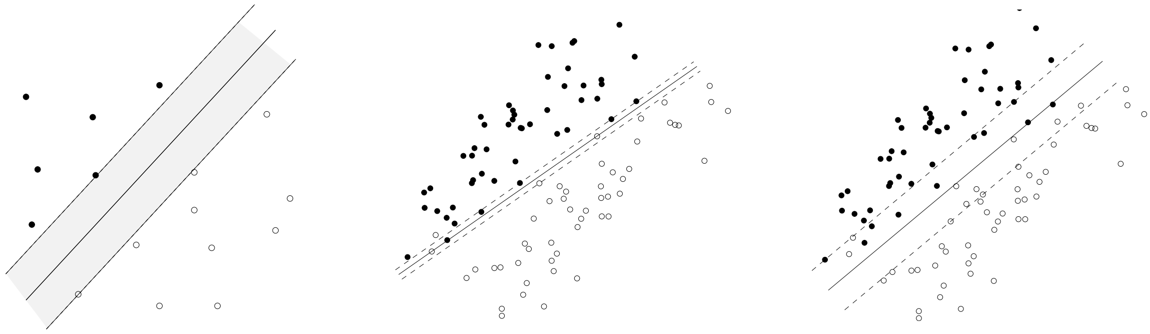


Le problème de séparation n'a pas toujours de solution ; ce qui nous amène à considérer trois situations, illustrés par les figures de la page suivante :

— (cas 1, figure de gauche) les deux ensembles sont bien séparables, et on cherche alors un $w$ qui les sépare "le plus possible" ;

— (cas 2, figure du milieu) les deux ensembles ne sont pas séparables, et on souhaite alors minimiser l'erreur de séparation faite par $w$ ;

— (cas 3, figure de droite) en pratique, on décide de faire un mélange des deux (minimiser l'erreur tout en séparant le plus possible).



**a)** Montrer que le problème de séparation décrit ci-dessus est "équivalent" au problème suivant

$$\text{(S)} \qquad \text{trouver } w \in \mathbb{R}^n \text{ tel que } \begin{cases} w^\top x_i \geqslant 1, & \text{pour tout } i \in I_+ \\ w^\top x_i \leqslant -1, & \text{pour tout } i \in I_- \end{cases}$$

au sens où : si $w$ est une solution de (S) alors c'est aussi une solution du problème initial, et inversement, si $w$ est une solution du problème initial, alors $\bar{w} = w/M$ est une solution de (S), avec

$$M_+ = \min_{i \in I_+} w^\top x_i\,, \qquad M_- = \max_{i \in I_-} w^\top x_i \qquad \text{et} \qquad M = \min\{M_+, -M_-\} \quad (>0).$$

Cet exercice étudit la résolution de ce problème (S) dans les trois cas illustrés ci-dessus. Commençons par le premier cas, où "séparer le plus possible" consiste à maximiser la distance entre les deux hyperplans $H_+(w) = \{x \in \mathbb{R}^n : \; w^\top x = 1\}$ et $H_-(w) = \{x \in \mathbb{R}^n : \; w^\top x = -1\}$. À $w$ (non-nul) fixé, cette distance est définie par

$$\operatorname{dist}(H_+(w), H_-(w)) \;=\; \begin{cases} \min & \frac{1}{2}\|x_+ - x_-\|^2 \\ & x_+ \in H_+(w), \quad x_- \in H_-(w). \end{cases}$$

**b)** Soit $H(w) = \{x \in \mathbb{R}^n \colon w^\top x = 0\}$ l'hyperplan orthogonal à $w$. Montrer que $H_+(w) = w/\|w\|^2 + H(w)$ (et de même $H_-(w) = -w/\|w\|^2 + H(w)$).

**c)** En déduire que $\operatorname{dist}(H_+(w), H_-(w)) = 2/\|w\|^2$.

**d)** Montrer que le problème dans le premier cas s'écrit ainsi

$$\begin{cases} \min & \|w\|^2/2 \\ & w^\top x_i \geqslant 1, \quad \text{pour tout } i \in I_+, \\ & w^\top x_i \leqslant -1, \quad \text{pour tout } i \in I_-. \end{cases}$$

Quelle est nature de ce problème ?

L'inconvénient de la modélisation sous la forme du problème précédent est qu'aucun $w$ ne va être rendu si les ensembles de points ne sont pas séparables. Pour traiter ce cas, on introduit l'erreur de séparation, comme la fonction en $w \in \mathbb{R}^n$ définie par

$$E(w) = \sum_{i=1}^{m} \max\left\{1 - y_i(w^\top x_i)\,, 0\right\} \qquad (\geqslant 0).$$

**e)** Montrer que $E$ est une fonction convexe.

**f)** Montrer que $E(w) = 0$ si et seulement si $w$ sépare bien les points (c'est-à-dire est solution de (S)).

On se place pour la fin de l'exercice dans le (troisième) cas où l'on veut à la fois minimiser erreur de séparation et maximiser la distance. Avec une constante $C > 0$ pour équilibrer les deux objectifs, on considère alors le problème d'optimisation convexe suivant :

$$\text{(P)} \quad \min_{w \in \mathbb{R}^n} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} \max\left\{1 - y_i(w^\top x_i), 0\right\}.$$

**g)** Montrer qu'il existe une unique solution à (P).

**h)** En introduisant $m$ variables supplémentaires $z_i \geqslant \max\{1 - y_i(w^\top x_i), 0\}$, montrer que (P) s'écrit comme un problème quadratique (convexe) en $n + m$ variables.

La nature du problème (P) est donc simple, mais en pratique sa résolution numérique est coûteuse, car $n$ est souvent très grand. Étudions alors une approche par dualité.

**i)** Écrire le problème quadratique de la question précédente (appelé "problème primal" dans la suite de l'exercice) sous la forme du cours gardant la contrainte $z \geqslant 0$ et dualisant les contraintes $1 - y_i(w^\top x_i) - z_i \leqslant 0$ (pour $i = 1, \ldots, m$). Montrer que lagrangien associé peut s'écrire, pour $(w, z) \in \mathbb{R}^n \times (\mathbb{R}_+)^m$ et $\mu \in (\mathbb{R}_+)^m$, comme

$$L((w,z);\mu) = \frac{1}{2}\left\|\sum_{i=1}^m \mu_i y_i x_i\right\|^2 - \mu^\top e - \frac{1}{2}\left\|w - \sum_{i=1}^m \mu_i y_i x_i\right\|^2 + z^\top(\mu - Ce)$$

où $e$ est le vecteur de $\mathbb{R}^m$ composé que de 1 (en scilab : `e=ones(m,1)`).

**j)** Fixons $\mu \geqslant 0$. Montrer que, si une composante de $\mu$ est telle que $\mu_j > C$, alors le sup du lagrangien $L((\cdot,\cdot);\mu)$ sur $\mathbb{R}^n \times (\mathbb{R}_+)^m$ vaut $+\infty$. Montrer que, si $0 \leqslant \mu \leqslant C$, alors le maximum du lagrangien est atteint pour $w = \sum_{i=1}^m \mu_i y_i x_i$, et tout $z \in (\mathbb{R}_+)^m$ tel que

$$\text{pour tout } i = 1, \ldots, m \quad \begin{cases} \text{si } \mu_i < C, \text{ alors } z_i = 0, \\ \text{si } \mu_i = C, \text{ alors } z_i \geqslant 0. \end{cases}$$

Donner l'expression de la fonction duale associée $\theta(\mu)$.

**k)** Montrer que le problème dual s'écrit comme le problème quadratique dans $\mathbb{R}^m$

$$\text{(D)} \quad \begin{cases} \min \ \frac{1}{2}\mu^\top Q\mu - \mu^\top e \\ 0 \leqslant \mu \leqslant Ce \end{cases}$$

où $Q$ est la matrice (carrée de taille $m$) définie par ses coefficients $Q_{ij} = y_i y_j {x_i}^\top x_j$.

**l)** Montrer que ce problème est convexe, qu'il en existe une solution, et que, si les vecteurs $x_1, \ldots, x_m$ de $\mathbb{R}^n$ sont indépendants, cette solution est unique.

**m)** Écrire sur papier les commandes scilab qui permettent de résoudre ce problème en appelant la fonction `qld` (dont l'aide scilab est donnée à la fin de ce sujet). Prendre, comme données, le réel $C > 0$, le vecteur $y \in \{-1, 1\}^m$ et la matrice $K \in \mathbb{R}^{m\times m}$ des produits scalaires ($K_{ij} = {x_i}^\top x_j$).

Le problème dual a ainsi les mêmes propriétes théoriques que le problème primal, mais il est bien plus facile à résoudre car sa taille ne dépend que de $m$, la taille de l'échantillon, et plus de $n$, la dimension du problème initial. De plus, les $x_i \in \mathbb{R}^n$ ne sont pas manipulés explicitement, mais uniquement via les produits scalaires ${x_i}^\top x_j$ définissant $K$ (c'est ce qu'on appelle le "kernel trick"). Nous allons maintenant montrer que cette approche duale donne aussi une solution primale.

**n)** Écrire les conditions d'optimalité (de KKT) du problème dual (D). Montrer qu'elles s'écrivent

$$\text{pour tout } i = 1, \ldots, m \qquad \begin{cases} \text{si } 0 < \mu_i < C, \text{ alors } 1 - \sum_{j=1}^m Q_{ij}\mu_j = 0 \\ \text{si } \mu_i = 0, \text{ alors } 1 - \sum_{j=1}^m Q_{ij}\mu_j \leqslant 0 \\ \text{si } \mu_i = C, \text{ alors } 1 - \sum_{j=1}^m Q_{ij}\mu_j \geqslant 0. \end{cases}$$

**o)** Notons $\mu^\star$ la solution optimale duale, et définissons

$$w^\star = \sum_{j=1}^m \mu_j^\star y_j x_j \qquad \text{et} \qquad z^\star = \left(\max\left\{0, 1 - \sum_{j=1}^m Q_{ij}\mu_j^\star\right\}\right)_{i=1,\ldots,m}.$$

Montrer que $(w^\star, z^\star)$ est réalisable pour le problème primal, et que de plus $\mu_i^\star(1 - y_i({w^\star}^\top x_i) - z_i^\star) = 0$ pour tout $i = 1, \ldots, m$.

**p)** En déduire que $w(\mu^\star)$ est la solution optimale de (P).

qld - linear quadratic programming solver

Calling Sequence

```
[x,lagr]=qld(Q,p,C,b,ci,cs,me [,tol])
[x,lagr,info]=qld(Q,p,C,b,ci,cs,me [,tol])
```

Parameters

Q: real positive definite symmetric matrix (dimension n x n).
p: real (column) vector (dimension  n)
C: real matrix (dimension  (me + md) x n)
b: RHS column vector (dimension  (me + md))

ci: column vector of lower-bounds (dimension n).
If there are no lower bound constraints, put ci = [].
If some components of x are bounded from below, set the other (unconstrained) values of ci to a very large negative number (e.g. ci(j) = -number_properties('huge').

cs: column vector of upper-bounds. (Same remarks as above).
me: number of equality constraints (i.e. C(1:me,:)*x = b(1:me))
tol:Floatting point number, required précision.

x: optimal solution found.
lagr: vector of Lagrange multipliers.
If lower and upper-bounds ci,cs are provided, lagr has n + me + md components and lagr(1:n) is the Lagrange vector associated with the bound constraints and lagr (n+1 : n + me + md) is the Lagrange vector associated with the linear constraints.
If no bounds are provided, lagr has only me + md  components.

info: integer, return the execution status instead of sending errors.
info==1 : Too many iterations needed
info==2 : Accuracy insufficient to statisfy convergence criterion
info==5 : Length of working array is too short
info==10: The constraints are inconsistent

Description

```
Minimize  0.5*x'*Q*x + p'*x
under the constraints
 C(j,:) x = b(j),  j=1,...,me
 C(j,:) x <= b(j), j=me+1,...,me+md
 ci <= x <= cs
```