

## TP 4 : OPTIMISATION CONVEXE

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, 15 OCTOBRE 2015

Jean-Baptiste Alayrac  
jean-baptiste.alayrac@inria.fr

RÉSUMÉ. Dans ce TP on va chercher à implémenter certaines notions vues en cours concernant l'optimisation convexe. Le premier exercice cherche à montrer que les conditions de KKT permettent souvent de donner une interprétation aux variables duales. Le deuxième a pour but de montrer que la vitesse de convergence d'un algorithme est considérablement influencée par certaines caractéristiques du problème. Enfin, on implémentera une méthode de Newton dans un dernier problème.

**Le code que vous serez emmené à faire pourra resservir dans la suite, conservez en une trace.**

**Remarque préliminaire :** Les vitesses de convergence des algorithmes de se représentent généralement sur un graphe logarithmique, il faut donc penser à utiliser les fonctions `semilogy`, `loglog` de Matlab/Octave...

### 1. EXERCICE 1 : DE L'INTÉRÊT DE LA DUALITÉ

Soit  $\alpha_1, \dots, \alpha_n \in \mathbb{R}_+^*$ , on considère le problème suivant :

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & - \sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{t.q} & \quad \mathbf{1}^T x = 1 \\ & \quad x \geq 0 \end{aligned}$$

1) En écrivant le Lagrangien, dérivez un dual à ce problème (on introduira des variables duales  $\lambda_i$  pour les contraintes d'inégalités, et  $\nu$  pour celle d'égalité).

2) En admettant qu'elles sont vérifiées à l'optimum, écrivez les conditions de KKT. Discutez suivant la valeur possible du multiplicateur  $\nu$  de la valeur des autres variables.

**Question/Remarque.** *Souvent, les variables duales et les conditions KKT peuvent être interprétées de manière "physique". Ici c'est le cas. On appelle le problème précédent, le problème du "remplissage par l'eau". En effet, il est usuel d'utiliser la discussion de la question précédente et de considérer  $n$  unités de terrain, chacune ayant une altitude  $\alpha_i$ , on cherche à inonder le terrain à un niveau uniforme avec une quantité unitaire de liquide. La variable duale  $1/\nu$  correspond au niveau final atteint par le liquide sur le terrain complet. Un dessin aide à comprendre...*

### 2. EXERCICE 2 : CONVERGENCE LINÉAIRE D'UN ALGORITHME DE DESCENTE DE GRADIENT

Implémentons l'algorithme de la descente de gradient avec un pas constant  $\gamma$  sur un problème simple : celui de la régression ridge. On rappelle qu'il s'agit du problème de minimisation du risque quadratique régularisé :

$$\min_{w \in \mathbb{R}^p} \frac{1}{2n} \|y - Xw\|_2^2 + \frac{\lambda}{2} \|w\|_2^2$$

- 3) Commencer par générer de manière aléatoire une matrice de design  $X \in \mathbb{R}^{n \times p}$  de taille  $n = 50$  dont chaque ligne  $x_i$  soit dans  $[0, 1]^p$  (on se place dans le cas  $p > n$ , par exemple  $p = 60$ ) et un vecteur de réponses  $Y \in \mathbb{R}^n$ .
- 4) Rappeler l'expression de l'estimateur de la régression ridge (annulez le gradient matriciel comme dans le premier cours).
- 5) On va maintenant illustrer la convergence d'une méthode de descente de gradient à pas constant vers cet optimum.
- a) La fonction que l'on minimise est elle strictement convexe ? Fortement convexe ? Si oui préciser la constante associée.
- b) Quel est le pas constant classique qui assure à une descente de gradient de converger ?
- c) Implémenter cette méthode de descente de gradient pour trouver le minimum et le vecteur réalisant ce minimum.
- d) Représenter graphiquement la vitesse de convergence.
- 6) Calculer le pas  $\gamma$  optimal de la descente de gradient (le pas qui fait le plus descendre la fonction objectif dans l'opposé de la direction du gradient).
- 7) Implémenter l'algorithme de minimisation associé. Comparer sa vitesse de convergence avec l'algorithme à pas constant.
- 8) Que se passe-t-il computationnellement si le paramètre de régularisation  $\lambda$  tend vers 0 ?

### 3. PROBLÈME : CENTRE ANALYTIQUE D'INÉGALITÉS PAR MÉTHODE DE NEWTON

Le but de ce problème est d'implémenter une méthode de Newton dans le cas d'un problème d'optimisation non-contraint sur des vecteurs de  $\mathbb{R}^d$ . Etant donné des réels  $b_1, \dots, b_n$  et des vecteurs  $a_1, \dots, a_n \in \mathbb{R}^d$ , on considère le problème suivant :

$$\min_{x \in \mathbb{R}^d} - \sum_{i=1}^n \log(b_i - a_i^T x)$$

*La résolution de ce problème s'appelle la recherche du centre analytique défini par les inégalités  $a_i^T x \leq b_i$*

9) Implémenter la résolution de ce problème par méthode de Newton. Vous penserez à ajuster le pas par backtracking line search. Essayez de représenter graphiquement l'évolution de l'erreur commise  $\|f_t - f^*\|_2$  en fonction du nombre d'itérations de l'algorithme (on pourra pour cela utiliser une méthode très simple : calculer la valeur de l'objectif au bout d'un très grand nombre d'itération et prendre cette valeur comme approximation de la vraie valeur de la fonction à l'optimum  $f^*$ ).

*L'optimisation de ce problème est une première étape vers des méthodes très largement utilisées à l'heure actuelle dans le domaine de l'optimisation non contrainte : les méthodes de point intérieur. Elles sont traitées in extenso dans le livre de Boyd et Vandenberghe servant de référence à cette partie du cours.*