TP 4 : OPTIMISATION CONVEXE

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, PRINTEMPS 2013

Rémi Lajugie
remi.lajugie@ens.fr

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

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

Remarque préliminaire : Pour représenter la vitesse de convergence d'un algorithme graphiquement, il vaut mieux le faire sur un graphe logarithmique...

1. Exercice 1 : Convergence linéaire d'un algorithme de descente de gradient

Pour commencer on va implémenter la méthode d'optimisation la plus simple : la descente de gradient avec un pas constant γ sur un problème simple : celui de la régression ridge dans le plan, qui correspond au problème de minimisation du risque quadratique régularisé (x correspond au vecteur augmenté avec des 1) :

$$\min_{w \in \mathbb{R}^2} \|y - w^T x\|_2^2 + \lambda \|w\|_2^2$$

1) Commencez par générer $n = 50$ données i.i.d (X_i, Y_i) avec X_i suivant une loi uniforme sur $[0,1]^p$ $(p > n)$ et Y_i générés par exemple suivant une loi normale.

2) Rappelez l'expression de l'estimateur de la régression ridge (annulez le gradient matriciel comme dans le premier cours).

3) On va maintenant représenter la convergence d'une méthode de descente de gradient vers cet optimum.

a) Reprenez le gradient matriciel de la fonctionnelle à minimiser. Est elle stritement convexe ? Fortement convexe ? Si oui précisez la constante associée.

b) Quel est le pas constant classique associé à une descente de gradient ?

c) Implémentez cette méthode de descente de gradient pour trouver le minimum et le vecteur réalisant ce minimum.

d) Représentez graphiquement la vitesse de convergence.

e) Que se passe-t-il si le paramètre de régularisation λ tend vers 0 ?

2. Exercice : De l'intérêt de la dualité

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

$$\begin{array}{cc} \min\limits_{x \in \mathbb{R}^n} & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ t.q & \mathbf{1}^T x = 1 \\ & x \geq 0 \end{array}$$

4) En écrivant le Lagrangien, dérivez un dual à ce problème (on introduira des variables duales λ_i pour les contraintes d'inégalités, et ν pour celle d'égalité).

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

6) **BONUS** : Souvent, les variables duales et les conditions KKT peuvent être interprétées de manière "physique". Ici c'est le cas. Il est usuel d'utiliser la discussion de la question précédente et de considérer n unités de terrain, chacune ayant une altitude α_i, on cherche à innonder 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. Essayez de compléter et visualiser cette interprétation en vous aidant d'un dessin.

7) Implémentez une méthode simple vue en cours pour résoudre ce problème. Le problème dual est il facile à résoudre ? Si oui, en utilisant la dualité forte, vous calculerez à chaque itération une borne supérieure sur l'écart de la valeur de la fonction évaluée au point courant de l'algorithme à l'optimum.

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 :

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

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

8) Implémentez 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.