Outils d'optimisation pour les sciences des données et de la décision

M2 MIAGE, 2023-2024

Contrôle continu (version du 6 février 2024)

Instructions

- La version la plus récente du projet (mise à jour en cas de changement) est disponible à l'adresse **https://www.lamsade.dauphine.fr/∼croyer/ensdocs/ODD/ProjODD.pdf**.
- Le travail décrit ci-dessous est à rédiger et remettre individuellement.
- Les réponses devront être envoyées au format PDF à `clement.royer@lamsade.dauphine.fr`. Le nom du fichier PDF devra clairement indiquer les nom et prénom de l'étudiant(e).
- La date limite pour transmettre les réponses est le **18 février 2024 AOE** (Anywhere On Earth).

Notations et remarques préliminaires

- Les dimensions des vecteurs ou matrices seront toujours supposées supérieures ou égales à 1.
- La notation $\|\cdot\|$ désignera la norme euclidienne.

Introduction

Lors du cours, nous avons essentiellement traité de problèmes sans contraintes explicites sur les variables de décision (à l'exception notable du cadre du dernier chapitre). Il est cependant très fréquent en pratique que les problèmes d'optimisation à résoudre comportent des contraintes, qu'il faut alors intégrer dans la définition d'une solution au problème, ainsi que dans le développement d'algorithmes pour résoudre numériquement des problèmes avec contraintes. Le but de ce projet est d'adapter certaines des méthodes vues en cours pour prendre en compte la présence de contraintes. Dans ce projet, on s'intéresse donc à des problèmes d'optimisation de la forme

$$\operatorname*{minimiser}_{\boldsymbol{w}\in\mathbb{R}^d} f(\boldsymbol{w}) \quad \text{sous contraintes } \boldsymbol{w}\in\mathcal{F}, \tag{1}$$

où f est une fonction de classe $\mathcal{C}^1$ et $\mathcal{F}$ est un ensemble convexe inclus dans $\mathbb{R}^d$.

Partie 1 : Problème de projection

Les méthodes que nous étudions ici reposent sur le principe de la projection sur l'ensemble convexe $\mathcal{F}$. Pour tout point $\boldsymbol{w} \in \mathbb{R}^d$, on définit la *projection de $\boldsymbol{w}$ sur $\mathcal{F}$* comme la solution du problème

$$\operatorname*{minimiser}_{\boldsymbol{u} \in \mathbb{R}^d} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{w}\|^2 \quad \text{sous contraintes} \quad \boldsymbol{u} \in \mathcal{F}. \tag{2}$$

On notera cette projection $P_{\mathcal{F}}[\boldsymbol{w}]$.

Question 1 *Justifier que la fonction objectif du problème (2) est fortement convexe.*

Question 2 *On suppose pour cette question que $\mathcal{F} = \mathbb{R}^d$. Montrer que le problème (2) possède une unique solution dans ce cas, et donner cette solution.*

Plus généralement, on peut montrer que le problème (2) possède toujours une unique solution lorsque l'ensemble convexe $\mathcal{F}$ vérifie certaines propriétés en plus de la convexité.[1]

Partie 2 : Solutions du problème

On rappelle qu'une solution du problème (1) est un vecteur $\boldsymbol{w}^* \in \mathbb{R}^d$ tel que

$$\boldsymbol{w}^* \in \mathcal{F} \quad \text{et} \quad f(\boldsymbol{w}^*) \leq f(\boldsymbol{w}) \quad \forall \boldsymbol{w} \in \mathcal{F} \tag{3}$$

et qu'une solution locale du problème (1) $\boldsymbol{w}^* \in \mathbb{R}^d$ est telle que

$$\boldsymbol{w}^* \in \mathcal{F} \quad \text{et} \quad \exists \epsilon > 0, f(\boldsymbol{w}^*) \leq f(\boldsymbol{w}) \quad \forall \boldsymbol{w} \in \mathcal{F}, \|\boldsymbol{w} - \boldsymbol{w}^*\| \leq \epsilon. \tag{4}$$

Enfin, sous les bonnes hypothèses sur $\mathcal{F}$, on peut établir une condition nécessaire d'optimalité pour le problème. Si $\bar{\boldsymbol{w}} \in \mathcal{F}$ est une solution locale du problème, alors

$$\|\bar{\boldsymbol{w}} - P_{\mathcal{F}}[\bar{\boldsymbol{w}} - \nabla f(\bar{\boldsymbol{w}})]\| = 0. \tag{5}$$

Il s'agit d'une implication et non d'une équivalence. On dira ainsi qu'un point $\bar{\boldsymbol{w}} \in \mathcal{F}$ vérifiant (5) est un point stationnaire pour le problème (1).

Question 3 *Dans le cas $\mathcal{F} = \mathbb{R}^d$, à quel concept vu en cours la condition (5) correspond-elle ?*

Dans le cas général, une solution du problème avec contraintes n'est en général pas une solution du problème sans contraintes.

Question 4 *Supposons que $d = 1$, $f(w) = \frac{1}{2}w^2$ et $\mathcal{F} = \{w \in \mathbb{R} \mid w \geq 1\}$. On rappelle que f est (notamment) de classe $\mathcal{C}^1$ avec $\nabla f(w) = w$ pour tout $w \in \mathbb{R}$.*

a) Donner une formule explicite pour le calcul de $P_{\mathcal{F}}[w]$.

b) En déduire que le point $w^* = 1$ est un minimum du problème (1) pour ces choix spécifiques de f et $\mathcal{F}$.

c) Justifier que w^* n'est en revanche pas un minimum du problème minimiser$_{w \in \mathbb{R}} f(w)$. Est-ce un point stationnaire de ce problème ?

[1] On ne mentionne pas ces propriétés ici par souci de simplicité. Celles-ci seront satisfaites pour les exemples qui nous intéresseront.

Partie 3 : Algorithme de gradient projeté

L'algorithme 1 décrit une version de base de l'algorithme du gradient projeté. À chaque itération, on définit un déplacement à partir d'un point auxiliaire qui représente la projection d'un pas dans la direction opposée au gradient.

Algorithm 1: Algorithme du gradient projeté pour le problème (1).

1. **Initialisation**: Choisir $\boldsymbol{w}_0 \in \mathcal{F}$.
2. **Pour** $k = 0, 1, \ldots$
   1. Calculer le gradient $\nabla f(\boldsymbol{w}_k)$ et $\bar{\boldsymbol{w}}_k = P_{\mathcal{F}}\left[\boldsymbol{w}_k - \nabla f(\boldsymbol{w}_k)\right]$.
   2. Déterminer une taille de pas $\alpha_k > 0$ telle que $\boldsymbol{w}_k + \alpha_k\left(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right) \in \mathcal{F}$.
   3. Poser $\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \alpha_k\left(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right)$.
3. **FinPour**

Question 5 *Supposons que $\boldsymbol{w}_k \in \mathcal{F}$ soit un point stationnaire du problème (1). Montrer que l'on a alors $\boldsymbol{w}_{k+1} = \boldsymbol{w}_k$.*

Question 6 *Supposons qu'on ne dispose pas d'un point initial vérifiant les contraintes. Partant d'un vecteur $\boldsymbol{w}_{-1} \notin \mathcal{F}$, proposer une manière de calculer un vecteur $\boldsymbol{w}_0 \in \mathcal{F}$ proche de $\boldsymbol{w}_{-1}$.*

Pour analyser l'algorithme 1, on se place dans le même cadre que celui de la descente de gradient, et on suppose que la fonction f est $\mathcal{C}_L^{1,1}$. On sait alors que l'on a notamment

$$f\left(\boldsymbol{w}_k + \alpha\left(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right)\right) \le f(\boldsymbol{w}_k) + \alpha \nabla f(\boldsymbol{w}_k)^{\mathrm{T}}\left(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right) + \frac{L}{2}\alpha^2\left\|\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right\|^2 \tag{6}$$

pour tout itéré $\boldsymbol{w}_k$ considéré dans l'algorithme 1 et tout $\alpha > 0$. De plus, on peut montrer que le couple $(\boldsymbol{w}_k, \bar{\boldsymbol{w}}_k)$ calculé à l'itération k de l'algorithme vérifie

$$\nabla f(\boldsymbol{w}_k)^{\mathrm{T}}(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k) \le -\|\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\|^2. \tag{7}$$

Question 7 *En utilisant les résultats du paragraphe précédent, justifier que le choix $\alpha = \frac{1}{L}$ dans l'équation (6) garantit $f\left(\boldsymbol{w}_k + \alpha\left(\bar{\boldsymbol{w}}_k - \boldsymbol{w}_k\right)\right) < f(\boldsymbol{w}_k)$ si $\boldsymbol{w}_k$ n'est pas un point stationnaire. Sous quelle condition cette valeur est-elle utilisable dans l'algorithme 1 ?*

Partie 4 : Lien avec le gradient proximal

Dans cette partie, on s'intéresse au lien entre l'algorithme du gradient projeté et celui du gradient proximal vu en cours. Partant du problème (1), on considère le problème régularisé

$$\underset{\boldsymbol{w} \in \mathbb{R}^d}{\text{minimiser}}\ f(\boldsymbol{w}) + \lambda\, \delta_{\mathcal{F}}(\boldsymbol{w}), \tag{8}$$

où $\lambda > 0$ et $\delta_{\mathcal{F}}$ est la fonction caractéristique de $\mathcal{F}$, définie par

$$\delta_{\mathcal{F}}(\boldsymbol{w}) := \begin{cases} 0 & \text{si } \boldsymbol{w} \in \mathcal{F} \\ +\infty & \text{sinon.} \end{cases}$$

Cette fonction est non lisse mais peut être considérée comme convexe dans la mesure où l'ensemble $\mathcal{F}$ est convexe. On peut donc envisager d'appliquer l'algorithme de gradient proximal au problème (8), ce qui conduit à des sous-problèmes de la forme

$$\boldsymbol{w}_{k+1} \in \operatorname*{argmin}_{\boldsymbol{w} \in \mathbb{R}^d} \left\{ f(\boldsymbol{w}_k) + \nabla f(\boldsymbol{w}_k)^{\mathrm{T}}(\boldsymbol{w} - \boldsymbol{w}_k) + \frac{1}{2\beta_k}\|\boldsymbol{w} - \boldsymbol{w}_k\|^2 + \lambda \delta_{\mathcal{F}}(\boldsymbol{w}) \right\}, \tag{9}$$

où $\beta_k > 0$. Sous les bonnes hypothèses sur $\mathcal{F}$, on peut montrer que la solution du sous-problème (9) est unique et donnée par

$$\boldsymbol{w}_{k+1} = P_{\mathcal{F}}\left[\boldsymbol{w}_k - \beta_k \nabla f(\boldsymbol{w}_k)\right]. \tag{10}$$

Question 8 *Sous quelles conditions l'itération (10) est-elle identique à celle de l'algorithme 1 ?*

Question 9 *On reprend l'exemple $d = 1$, $f(w) = \frac{1}{2}w^2$ et $\mathcal{F} = \{w \in \mathbb{R} \mid w \geq 1\}$ déjà considéré à la question 4. En supposant que $w_0 = 2$, comparer les deux itérations suivantes :*

1. *Itération de l'algorithme 1 avec $\alpha_k = \frac{1}{2}$.*
2. *Itération (10) avec $\beta_k = \frac{1}{2}$.*

Partie 5 : Pénalisation

Dans cette dernière partie, plutôt que de pénaliser les vecteurs n'appartenant pas à $\mathcal{F}$ par une valeur infinie, on considère une approche dite de pénalisation

$$\operatorname*{minimiser}_{\boldsymbol{w} \in \mathbb{R}^d} f(\boldsymbol{w}) + \lambda\, d(\boldsymbol{w}; \mathcal{F})^2, \tag{11}$$

où $\lambda > 0$ et $d(\cdot; \mathcal{F})$ est la fonction distance d'un vecteur à un ensemble, définie par

$$\forall \boldsymbol{w} \in \mathbb{R}^d, \qquad d(\boldsymbol{w}; \mathcal{F}) := \min_{\boldsymbol{u} \in \mathbb{R}^d} \left\{ \|\boldsymbol{u} - \boldsymbol{w}\| \text{ sous contraintes } \boldsymbol{u} \in \mathcal{F} \right\}. \tag{12}$$

Question 10 *Quel est le lien entre la fonction distance et le problème de projection (2) ?*

Un algorithme de gradient proximal appliqué au problème (11) consiste alors en l'itération :

$$\boldsymbol{w}_{k+1} \in \operatorname*{argmin}_{\boldsymbol{w} \in \mathbb{R}^d} \left\{ f(\boldsymbol{w}_k) + \nabla f(\boldsymbol{w}_k)^{\mathrm{T}}(\boldsymbol{w} - \boldsymbol{w}_k) + \frac{1}{2\beta_k}\|\boldsymbol{w} - \boldsymbol{w}_k\|^2 + \lambda\, d(\boldsymbol{w}; \mathcal{F})^2 \right\}. \tag{13}$$

Question 11

a) Quel peut être l'inconvénient pratique de l'itération (13) ?

b) Sachant que la fonction $\boldsymbol{w} \mapsto d(\boldsymbol{w}; \mathcal{F})$ est convexe, proposer un algorithme applicable à la résolution du sous-problème (13).