M2 Master Mathematics for data science and big data
Computational Statistics and Optimization I

Exam

Whithin exercises the questions are *not* independent, but, at each step, you may use the results of previous questions, even if you have not succeeded in proving them.

No documents and electronic devices are allowed

Exercice 1

(a) (1 point) Consider the problem

$$\min_{x\in\mathbb{R}} \quad (x-1)^2 + \lambda|x| \tag{1}$$

where $\lambda \geq 0$. Give a necessary and sufficient condition on λ to ensure that 0 is a minimizer of problem (1).

(b) (1 point) Solve the problem (find a minimizer and the corresponding value) for any $\lambda \geq 0$.

Exercice 2

(a) (0.5 points) Consider the problem

$$\min_{x\in\mathbb{R}} \quad (x-1)^2 \quad \text{such that} \quad x \leq a \tag{2}$$

where $a \in \mathbb{R}$. Do the KKT and Lagrangian saddle point theorems apply? Justify your answer.

(b) (0.5 points) Write the Lagrangian function for problem (2) and compute the dual function $\mathcal{D}(\lambda)$ ($\lambda \in \mathbb{R}$).

(c) (1 point) Write the Lagrangian dual problem and solve it (find a dual optimal point λ^* and compute the dual value d).

(d) (1 point) At this point, what can you say about the primal value p?

(e) (1 point) Using Question (c), show that the primal optimal x^* is solution of an unconstrained problem. Solve this problem (compute x^* and p)

(f) (1 point) Write the KKT condition for problem (2)). Using these conditions, show directly that $x^* = 1 \Rightarrow a \geq 1$.

Exercice 3

(a) (0.5 points) Recall the respective definitions of *proper function* and *lower semi-continuous* (l.s.c.) function.

(b) (0.5 points) What is an affine minorant of a function f? Give a relation between a function and its affine minorants. Detail precisely the conditions under which your statement is valid.

(c) (2 points) Let $f : \mathbb{R} \to (-\infty, \infty]$ be a proper, (l.s.c.), convex function. Let $g : \mathbb{R}^2 \to (-\infty, \infty]$ be defined by

$$g(x,y) = \begin{cases} y\, f(x/y) & \text{if } y > 0 \\ +\infty & \text{otherwise} \end{cases}$$

Show that g is convex (use question (b)).

(d) (3 points) Show that, $\forall(\lambda, \phi) \in \mathbb{R}^2$

$$g^*(\lambda, \phi) = \begin{cases} 0 & \text{if } f^*(\lambda) + \phi \leq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

(e) (3 points) Show that

$$g^{**}(x, y) = \begin{cases} yf(x/y) & \text{if } y > 0 \\ \sup_{\lambda : f^*(\lambda) < \infty} \lambda x & \text{if } y = 0 \\ +\infty & \text{if } y < 0. \end{cases}$$

Exercice 4

We consider the problem of multitask-regression, where the aim is to predict several labels $y^{(1)}, \dots, y^{(K)} \in \mathbb{R}$ at the same time, based on features $x_1, \dots, x_p \in \mathbb{R}$. We say that K is the number of tasks and p is the number of features. We have a training set containing n samples of labels and features, that we concatenate in an $n \times K$ matrix Y and an $n \times d$ matrix X. For a parameter $\Theta \in \mathbb{R}^{p \times K}$, we consider the multitask least-squares loss $\Theta \mapsto \frac{1}{2nK}\|Y - X\Theta\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm given by $\|Z\|_F = \sqrt{\sum_{i,j}(Z_{i,j})^2}$, which is the norm associated to the Euclidean inner product for matrices, given by $\langle \Theta, \Theta' \rangle = \sum_{j=1}^{p} \sum_{k=1}^{K} \Theta_{j,k}\Theta'_{j,k}$. We consider a procedure based on a penalization of this loss, namely

$$\hat{\Theta} \in \underset{\Theta \in \mathbb{R}^{p \times K}}{\operatorname{argmin}} \left\{ \frac{1}{2nK}\|Y - X\Theta\|_F^2 + \lambda\Omega(\Theta) \right\}, \tag{3}$$

where

- $Y = [Y_1, \cdots, Y_K] \in \mathbb{R}^{n \times K}$ contains the observed labels
- $X = [X_1, \cdots, X_p] \in \mathbb{R}^{n \times p}$ contains the features
- $\Theta = [\Theta_1, \cdots, \Theta_K] \in \mathbb{R}^{p \times K}$ is the parameter to be estimated.

In the following we consider the group-Lasso (ℓ_1/ℓ_2) penalization:

$$\Omega(\Theta) = \sum_{j=1}^{p} \|\Theta_{\bullet,j}\|_2,$$

where $\|\cdot\|_2$ is the ℓ_2-norm and $\Theta_{\bullet,j}$ is the j-th line of the matrix Θ.

(a) (1 point) We define $f(\Theta) = \frac{1}{2nK}\|Y - X\Theta\|_F^2$. Compute $\nabla f(\Theta)$.

(b) (1 point) Compute the Lipschitz constant L of ∇f.

(c) (3 points) Recall the definition of the proximal operator. Prove that

$$\partial\|x\|_2 = \begin{cases} \frac{x}{\|x\|_2} & \text{if } x \neq 0 \\ \{z : \|z\|_2 \leq 1\} & \text{if } x = 0. \end{cases}$$

Use this result together with the optimality condition of the proximal operator to compute the proximal operator of $\lambda\Omega$ of a matrix Θ. Explain with words what this operators does to Θ. Why this penalization seems pertinent?

(d) (2 points) Propose a proximal gradient descent algorithm to solve (3). Describe precisely the steps.

(e) (3 points) Give a definition of the duality gap. Compute this quantity for the multi-task problem (3), and deduce a stopping criterion based on the duality gap.