SD-TSIA 211
Optimization for Machine Learning
24 January 2020

Paper documents are allowed (lecture notes, exercises and books)
Electronic devices are forbidden

Exercise 1 (Dual of a quadratic program).
We consider the problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top Q x + c^\top x \qquad (1)$$
$$Ax \leq b$$

where Q is a symmetric positive definite matrix of size $n \times n$, $c \in \mathbb{R}^n$, A is a $m \times n$ matrix and $b \in \mathbb{R}^m$. The inequality $Ax \leq b$ means $(Ax)_j \leq b_j$ for all $j \in \{1, \ldots m\}$.

1. How many equality constraints does Problem (1) have ? How many inequality constraints ?

0 equality
m inequalities

2. Write the Lagrangian function $L(x, \varphi)$ associated to Problem (1).
3. Recall the definition of the dual problem.
4. Given $\varphi \in \mathbb{R}^m_+$, find $x(\varphi) \in \arg\min_x L(x, \varphi)$.

2 Lagrangian funct

$$L(x,\varphi) = \frac{1}{2} x^T Q x + c^T x + \varphi^T A x - \varphi^T b - \iota_{\mathbb{R}_+^m}(\varphi)$$

3 Dual problem : $\sup_{\varphi \in \mathbb{R}^m} D(\varphi)$

Where $D(\varphi) = \inf_{x \in \mathbb{R}^n} L(x,\varphi)$

4 $\arg\min_x L(x,\varphi) \Rightarrow x^*$ so that $L'(x^*,\varphi) = 0$

$$\Rightarrow L'(x^*,\varphi) = \frac{1}{2} \overbrace{(Q+Q^T)}^{2Q} x^* + c + A^T\varphi = 0 \qquad \frac{\partial}{\partial x}(x^T Q x) = (Q+Q^T)x$$

$$\Rightarrow \boxed{x^* = -Q^{-1}(c + A^T\varphi)} \rightsquigarrow x(\varphi) = -Q^{-1}(c + A^T\varphi)$$

$$\Rightarrow x^{*T} = -(c^T + \varphi^T A)\, Q^{-1}$$

5. Show that the dual of Problem (1) is a quadratic problem of the form

$$\max_{\varphi \in \mathbb{R}^m} \; -\frac{1}{2}\varphi^\top M \varphi + d^\top \varphi + a + \iota_{\mathbb{R}_+^m}(\qquad (2)$$
$$\varphi \geq 0$$

where you should give the expression of the matrix M, the vector d and the real number a.

6. Calculate $\operatorname{prox}_{\iota_{\mathbb{R}_+^m}}(x)$ where

$$\iota_{\mathbb{R}_+^m} : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$$
$$x \mapsto \begin{cases} 0 & \text{if } x_j \geq 0, \forall j \in \{1, \ldots, m\} \\ +\infty & \text{otherwise} \end{cases}$$

yi = max{0;xi} for i e {1,...,m}

Exercise 2 (ℓ_1 regression).
Basic results
We admit that for all x, the gradient descent method applied to a function F satisfies

$$F(x_k) - F(x) \leq \frac{L}{2k} \|x_0 - x\|_2^2$$

where L is the Lipschitz constant of the gradient of F and the result holds even if x is not a minimizer of F.

We recall that $\|x\|_1 = \sum_{j=1}^m |x_j|$, $\|x\|_2 = \sqrt{\sum_{j=1}^m |x_j|^2}$, $\|x\|_\infty = \max_{j=1}^m |x_j|$ and the ball of radius r for the norm $\|\cdot\|$ is $B_r = \{x \, : \, \|x\| \leq r\}$.

Setup
In this exercise we study the following ℓ_1 regression problem

$m \times n \cdot n \cdot 1$

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1 \tag{3}$$

where $A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix}$ is a $m \times n$ matrix and $b \in \mathbb{R}^m$. We admit that there exists at least a minimizer x^* to Problem 3. We are interested in a method to solve this problem, which is based on a technique called smoothing.
We fix $\mu > 0$ and we define

$$f(x) = \|Ax - b\|_1 = \sum_{j=1}^m |a_j^\top x - b_j|$$

$$f_\mu(x) = \max_{u \in [-1,1]^m} \langle u, Ax - b \rangle - \frac{\mu}{2} \|u\|_2^2$$

$$u_\mu(x) = \arg \max_{u \in [-1,1]^m} \langle u, Ax - b \rangle - \frac{\mu}{2} \|u\|_2^2 .$$

1. Is f convex? differentiable?

Convex because the norm 1 is convex
Not differentiable because f has a discontinuity in Ax=b

Phi(z) = |Z|
u need to have the same sign of z to Phi(z) have its maximum.

3. Show that $f(x) = \max_{u \in [-1,1]^m} \langle u, Ax - b \rangle$.

Max u in [-1,1] <u, Ax-b> can be written as abs(Ax - b), which is equal to f(x)

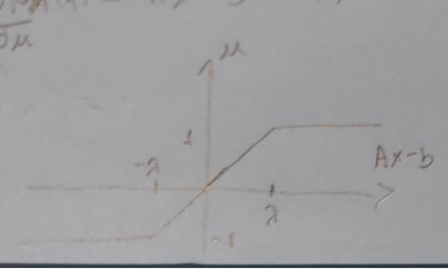
4. Why is there only one maximizer $u_\mu(x)$ to the problem defining $f_\mu(x)$?

The second derivative is -\mu. As \mu is always positive, it's second derivative is negative, then the function is STRICTLY concave, having an unique maximum.

A function being strictly concave does not mean it has a unique maximum (take f(x) = 1-e^(-x) for example) , so what is the answer, do you know? (Filipe)
Strongly concave should be enough though? Yes, strongly concave works.

$\boxed{4}$ $\frac{\partial f_\lambda'(x)}{\partial \mu} = Ax - b - \lambda \mu^* = 0$

$\underset{\in[-1,1]}{\mu^*} = \frac{1}{\lambda}(Ax-b)$



$$\Rightarrow \mu_i^* = \begin{cases} -1, & a_i^T x - b_i < -\lambda \\ \frac{1}{\lambda}(a_i^T x - b), & |a_i^T x - b_i| < \lambda \\ 1, & a_i^T x - b > \lambda \end{cases}$$

for $i = 1, 2, \ldots, n$

5. Show that for all $x \in \mathbb{R}^n$, $f_\mu(x) \leq f(x)$.

As we show that f(x) can be written as Max u in [-1,1] <u, Ax-b>, f\mu = f(x) - \mu/2 ||u||^2 (which is always positive), so the second term is always negative. Thus, we have that f\mu(x) <=f(x) (Filipe)

6. Show that for all $x \in \mathbb{R}^n$, $f_\mu(x) \geq f(x) - \mu D^2$, where D is a quantity to be determined.

$f_\mu(x) + \mu D^2 >= f(x)$ → $\mu D^2 >= \mu/2 \, ||u||^2$ → $D^2 >= ||u||^2/2$ (right?) Seems good!

7. Show that $u_\mu(x) = P_{B_\infty}(V)$ where P_{B_∞} is the projection onto the ball of the infinity norm with radius 1 and V is a quantity to determine as a function of μ, A, b and x.

Vi = sgn(Aix-bi-mu/2)