Fatma Moalla

Homework 3

Question 1

The lasso problem is the following:

$$min_w \frac{1}{2}\|Xw - y\|_2^2 + \lambda\|w\|_1$$

this problem is equivalent to the following problem if we introduce a latent variable z=Xw

$$min_w \frac{1}{2}\|z - y\|_2^2 + \lambda\|w\|_1$$

$$z = Xw$$

In order to derive the dual problem, we should compute the Lagrangian, let γ be the Lagrangian multiplier with only positive components

then

$$L(z, w, \gamma) = \frac{1}{2}\|z - y\|_2^2 + \gamma\lambda\|w\|_1 + \gamma^T(z - Xw)$$

The dual problem is therefore

$$max_\gamma[min_{z,w} L(z, w, \gamma)]$$

1. First step : minimize the Lagrangian over z and w :
   $L(\gamma) = min_{z,w} \frac{1}{2}\|z - y\|_2^2 + \gamma\lambda\|w\|_1 + \gamma^T(z - Xw) =$
   $min_z[\frac{1}{2}\|z - y\|_2^2 + \gamma^T z] + min_w[\gamma\lambda\|w\|_1 - \gamma^T Xw]$
   $= min_z F(z) + min_w G(w)$
   - We compute the gradient of F and set it to zero because F is differentiable: therefore we have $\frac{\partial F}{\partial z} = z - y + \gamma = 0$ then $z = y - \gamma$
   - By computing the dual norm of L^1
     $, min_w G(w) = -max_w\{\gamma^T Xw - \gamma\lambda\|w\|_1\} = \begin{cases} 0 & \text{if } \|X^T\gamma\|_\infty <= \lambda \\ -\infty & \text{else} \end{cases}$
2. Second, we get the following dual problem:

$$min_\gamma[\frac{1}{2}\gamma^T\gamma - \gamma^T y]$$

$$s.t$$

$$\|X^T\gamma\|_\infty <= \lambda$$

3. Third, we transform the dual problem into a QP problem by transforming the inequality constraint:

$$min_v[\frac{1}{2}v^Tv + y^Tv]$$
$$s.t$$
$$X^Tv <= \lambda\mathbf{1}_n$$
$$and$$
$$-X^Tv <= \lambda\mathbf{1}_n$$

with : $Q = \frac{1}{2}\mathbf{I}_n$ then **Q is Positive Semi Definite**, $P = y$, $A = [X, -X]^T$ and $b = \lambda\mathbf{1}_{2n}$

Question 2

In order to apply the **Newton Method** for the **centering step**, we need to compute the gradient and the hessien of the objective function f

1. In order to apply the newton method for the centering step, we need to compute the gradient and the hessien of the objective function f

   - Gradient computation

   $$\nabla(f)(v) = t(2Q.v^T + p) - \sum_i \frac{1}{(Av)_i - b_i}A_i$$

   - Hessien computation

   $$\nabla^2(f)(v) = tQ + \sum_i \frac{1}{((Av)_i - b_i)^2}A_i.A_i^T$$

2. in order to test the **centering step** and **barrier method**, we chose the following inputs:

   $A = random(10 * 10)$ matrix

   $Q = \frac{1}{2}\mathbf{I}_{10}$

   $p = random(10 * 1)$ vector

   $b = 10\mathbf{I}_{10}$

$v0 = random(10 * 1)$ vector

$\mu = 15$

$\epsilon = 10^{-3}$

The optimal v resulted from the barrier method is the following

$v = [[-1.00000408][-1.00000637][-1.00000642][-1.00000458]$
$[-1.00000799][-1.00000683][-1.00000798][-1.00000611][-1.00000652][-1.00000667]]$

3. We also checked that the inequality constraint is satisfied (detailed in the notebook).

Question 3

1. We selected randomly distributed inputs as following:

   $X = random(10 * 100)$ matrix

   $A = [X, -X]^T$

   $Q = \frac{1}{2}\mathbf{I}_1 0$

   $\lambda = 10$

   $p = 100 * random(10 * 1)$ vector

   $b = \lambda \mathbb{I}_{10}$

   $v0 = random(10 * 1)$ vector

   $\epsilon = 10^{-6}$

2. We applied the barrier method on the following list of μ values:

   $\mu = [2, 3, 5, 15, 50, 100, 200, 500]$

   The resulted graph is shown in Figure 1

3. since $v = y - z$ and $z = Xw$ we can deduce the influence of μ on w from its influence on z (here X is not a square matrix, therefore it is not invertible).

   The Graphic in Figure 2 shows this influence on the values of z (detailed on the notebook).

4. **Comments**

   - Figure 1 : The Best choice of μ would be greater or equal to 50, as it converges quickly and in **20 iterations or less** . As for the range of μ that we chose, the best $\boldsymbol{\mu = 500}$.
   - Figure 2: We know that $z = Xw = y - v$ and that the more mu increases, the larger is the step on minimization which means that convergences is quicker with larger μ which is shown in the graph below.

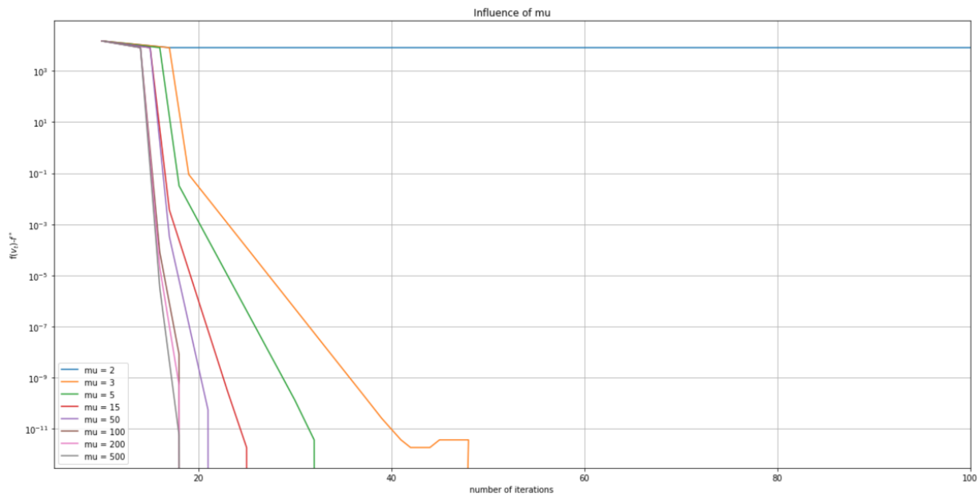

Figure 1: Influence of μ on the dual gap

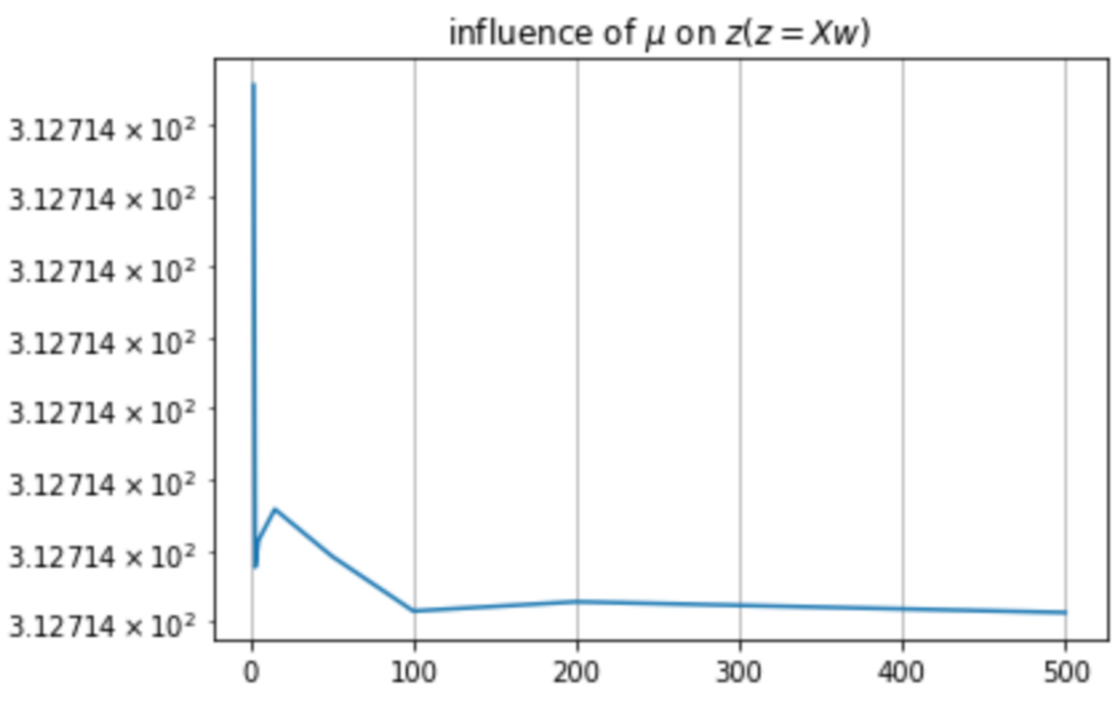

Figure 2: Influence of μ on z