Convex Optimization - Homework 3

Bastien DECHAMPS

November 25, 2019

Given $x_1, ..., x_n \in \mathbb{R}^d$ data vectors and $y_1, ..., y_n \in \mathbb{R}$ observations, we want to solve the following problem called LASSO:

$$\text{Min} \quad \frac{1}{2} \|Xw - y\|_2^2 + \lambda \|w\|_1 \tag{LASSO}$$

in the variable $w \in \mathbb{R}^d$, where $X = \left(x_1^\top, ..., x_n^\top\right) \in \mathbb{R}^{n \times d}$, $y = (y_1, ..., y_n) \in \mathbb{R}^n$ and $\lambda > 0$ is a regularization parameter.

1 Dual problem of LASSO

To compute the dual of (LASSO), we introduce a dummy variable $u = Xw - y$ to add an equality constraint. The problem is now equivalent to

$$\begin{array}{ll} \text{Min} & \frac{1}{2}\|u\|_2^2 + \lambda \|w\|_1 \\ \text{s.t.} & u = Xw - y \end{array}$$

The lagrangian $\mathcal{L}$ of this problem can be written:

$$\begin{aligned} \mathcal{L}(w, u, \nu) &= \frac{1}{2}\|u\|_2^2 + \lambda \|w\|_1 + \nu^\top (Xw - y - u) \\ &= \underbrace{\frac{1}{2}\|u\|_2^2 - \nu^\top u}_{F(u)} + \underbrace{\lambda \|w\|_1 + \nu^\top X w}_{G(w)} - \nu^\top y \end{aligned}$$

And thus we have

$$g(\nu) \doteq \inf_{w,u} \mathcal{L}(w, u, \nu) = \inf_u F(u) + \inf_w G(w) - \nu^\top y$$

We compute $\inf_u F(u)$ by solving $\nabla F(u) = u - \nu = 0$, thus the minimum is reached on $u = \nu$ and

$$\inf_u F(u) = -\frac{1}{2}\|\nu\|_2^2$$

We can rewrite $\inf_w G(w)$ using the conjugate fonction of $\|.\|_1$ denoted f^*:

$$\begin{aligned} \inf_w G(w) &= \inf_w \lambda \|w\|_1 + \nu^\top X w \\ &= -\lambda \sup_w \left[\left(-\frac{X^\top \nu}{\lambda} \right)^\top w - \|w\|_1 \right] \\ &= -\lambda f^* \left(-\frac{X^\top \nu}{\lambda} \right) \\ &= \left\{ \begin{array}{ll} 0 & \text{if } \|X^\top \nu\|_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{array} \right. \end{aligned}$$

Finally, the dual problem of (LASSO) is:

$$\begin{array}{ll} \text{Max} & -\frac{1}{2}\|\nu\|_2^2 - \nu^\top y \\ \text{s.t.} & \|X^\top \nu\|_\infty \leq \lambda \end{array}$$

which is equivalent to the Quadratic Problem:

$$\boxed{\begin{array}{ll} \text{Min} & \frac{1}{2}\nu^\top \nu + y^\top \nu \\ \text{s.t.} & A\nu \preceq \lambda \mathbb{1}_{2d} \end{array}} \tag{QP}$$

where

$$\mathbb{1}_k = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^k, \quad A = \begin{pmatrix} X^\top \\ -X^\top \end{pmatrix} \in \mathbb{R}^{2d \times n}$$

2 Experiments

The barrier method is implemented in Python in the file `barrier.py`. To run the experiments, type `python barrier.py` on a command line, provided that you have `numpy` and `scikit-learn` (for data generation) installed.

I used the parameters $\alpha = 0.25$ and $\beta = 0.9$ for backtracking line search, $t_0 = 0.1$ for the barrier method, and a precision $\varepsilon = 10^{-7}$. The values of $f(v_t) - f(v_T)$ are plotted on the figure 1, where T is the last iteration of the method.

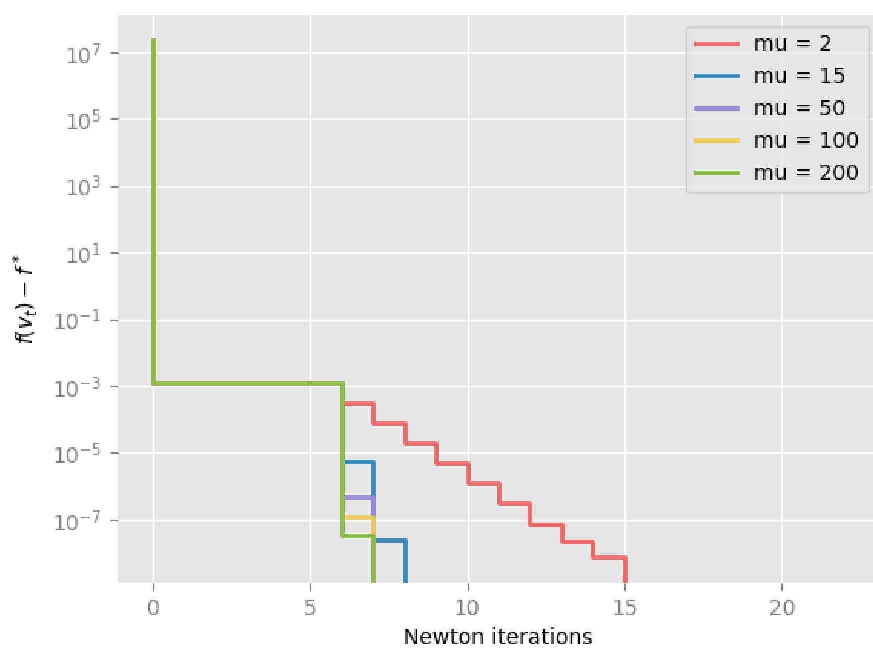

Figure 1: Convergence of the barrier method for different values of μ

For small values of μ, the barrier method takes more iterations to converge. However, we cannot really choose an optimal μ parameter among $\{15, 50, 100, 200\}$ only with this figure. I plotted the real duality gap on figure 2 by computing the primal problem objective through iterations. Indeed, by differenciating the lagrangian, we have:

$$Xw^* - Y - v^* = 0 \tag{1}$$

and thus, by a least square regression, we can obtain $w^* = \left(X^\top X\right)^{-1} X^\top (v^* + Y)$.

Finally, the value $\mu = 15$ seems to give the lowest duality gap.

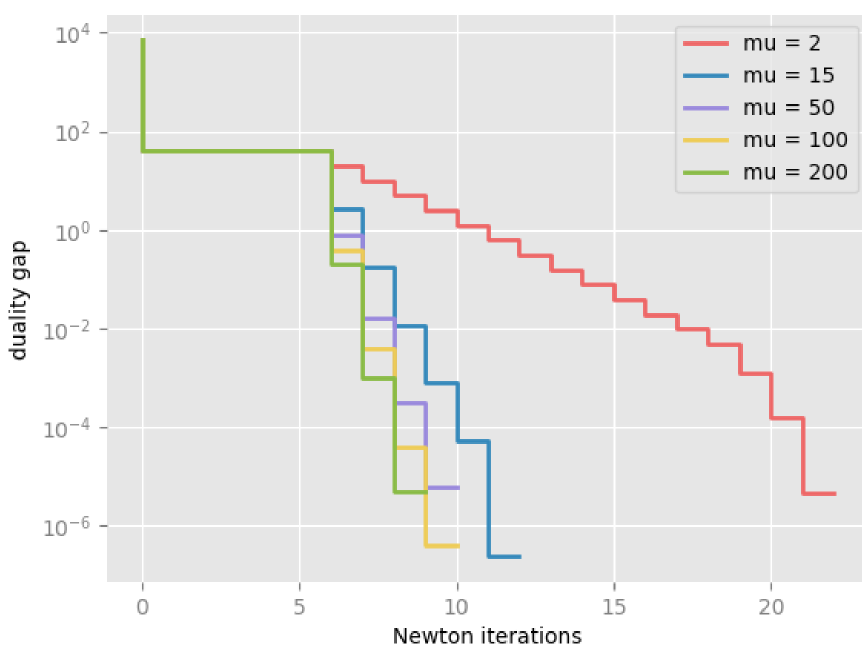

Figure 2: Duality gap evolution through Newton iterations for different values of μ