HW3 Report

Imen AYADI

November 2020

1 Question 1 : Dual of LASSO

In the previous homework, we have shown that for $u \in \mathbb{R}^n$,

$$\|\cdot\|_1^*(u) = \begin{cases} 0 & \text{if } \|u\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

The problem of LASSO is equivalent to:

$$\boxed{\begin{array}{l} \min_{w,z} \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 \\ \\ \text{s.t. } z = Xw - y \end{array}} \tag{1}$$

Let $w \in \mathbb{R}^d, z \in \mathbb{R}^n, \nu \in \mathbb{R}^n$. The Lagrangian function of the new problem Equation (1) is given by:

$$\begin{aligned} \mathcal{L}(w, z, \nu) &= \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 + \nu^T (z - (Xw - y)) \\ &= \frac{1}{2}\|z\|_2^2 + \nu^T z + \lambda\|w\|_1 - \left(X^T\nu\right)^T w + \nu^T y \end{aligned}$$

Then,noticing that the Lagrangian is a sum of separable functions of w and z, the dual function is given by:

$$\begin{aligned} g(\nu) &= \inf_{w,z} \mathcal{L}(w, z, \nu) \\ &= \nu^T y + \inf_z \left(\frac{1}{2}\|z\|_2^2 + \nu^T z\right) + \inf_w \left(\lambda\|w\|_1 - \left(X^T\nu\right)^T w\right) \end{aligned}$$

The function $\rho : z \mapsto \frac{1}{2}\|z\|_2^2 + \nu^T z$ is (strictly) convex and differentiable. Its gradient is given by $\nabla\rho(z) = z + \nu$ and we have $\nabla\rho(z) = 0$ iff $z = -\nu$. Hence, $-\nu = \text{argmin}_z \rho(z)$ and the minimum of ρ is $\rho(-\nu) = -\frac{1}{2}\|\nu\|_2^2$.
Furthermore, using the conjugate of $\|\cdot\|_1$, we get:

$$\inf_w \lambda\|w\|_1 - \left(X^T\nu\right)^T w = -\lambda \sup_w \left((\frac{1}{\lambda}X^T\nu)^T w - \|w\|_1 \right) = -\lambda\|\cdot\|_1^* \left(\frac{1}{\lambda}X^T\nu\right)$$

Thus,

$$g(\nu) = \begin{cases} \nu^T y - \frac{1}{2}\|\nu\|_2^2 & \text{if } \|\frac{1}{\lambda}X^T\nu\|_\infty \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

Hence, the dual problem of LASSO is:

$$\boxed{\begin{array}{l} \max_\nu -\frac{1}{2}\|\nu\|_2^2 + \nu^T y \\ \\ \text{s.t. } \|X^T\nu\|_\infty \leq \lambda \end{array}} \tag{2}$$

Let's reformulate the inequality constraint :

$$\begin{aligned} \|X^T\nu\|_\infty \leq \lambda \text{ iff } & \quad \forall i \in [\![1, d]\!], -\lambda \leq \left[X^T\nu\right]_i \leq \lambda \\ \text{iff } & \quad \forall i \in [\![1, d]\!], \left[X^T\nu\right]_i \leq \lambda \text{ and } \left[-X^T\nu\right]_i \leq \lambda \\ \text{iff } & \quad A\nu \preceq \lambda.\mathbf{1_{2d}} \end{aligned}$$

where $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix} \in \mathbb{R}^{2d\times n}$.
Hence, after plugging the reformulated constraint in the dual Equation (3), we get:

$$\boxed{\begin{array}{l} \min_\nu \nu^T Q\nu + p^T\nu \\ \\ \text{s.t. } A\nu \preceq b \end{array}} \tag{3}$$

where $Q = \frac{1}{2}I_n,\ p = -y,\ b = \lambda.\mathbf{1_{2d}}$ and $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix} \in \mathbb{R}^{2d\times n}$

2 Question 2 : Barrier Method

The goal of the centering step is to solve the unconstrained problem:

$$\min_{\nu} g_t(\nu) \triangleq t\left(\nu^T Q\nu + p^T\nu\right) - \sum_{i=1}^{2d} \log\left(b_i - [A\nu]_i\right)$$

We can write:

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_{2d}^T \end{pmatrix}$$

where a_i^T is the i^{th} row of A ; ($a_i \in \mathbb{R}^n$).
Let's compute the gradient and the Hessian matrix of the objective function:

$$\nabla g_t(\nu) = t\left(2Q\nu + p\right) + \sum_{i=1}^{2d} \frac{a_i}{b_i - a_i^T\nu} \tag{4}$$

and,

$$\nabla^2 g_t(\nu) = 2t\ Q + \sum_{i=1}^{2d} \frac{1}{(b_i - a_i^T\nu)^2} a_i a_i^T \tag{5}$$

For get more "beautiful" formulas to implement, we can define the matrix $\tilde{A}$ as:

$$\tilde{A} = \begin{pmatrix} \frac{a_1^T}{b_1 - a_1^T\nu} \\ \vdots \\ \frac{a_{2d}^T}{b_{2d} - a_{2d}^T\nu} \end{pmatrix}$$

Then,

$$\boxed{\nabla g_t(\nu) = t\left(2Q\nu + p\right) + \tilde{A}^T \cdot \mathbf{1_{2d}}} \tag{6}$$

and,

$$\boxed{\nabla^2 g_t(\nu) = 2t\ Q + \tilde{A}^T\tilde{A}} \tag{7}$$

For the code, see `HW3Ayadi.ipynb`.

3 Question 3 : Impact of μ

When we plot the precision criterion , we notice that when μ increases, the speed of the convergence increases.
I checked this fact for different starting points (different values of t_0) and I got similar curves. I also tried different

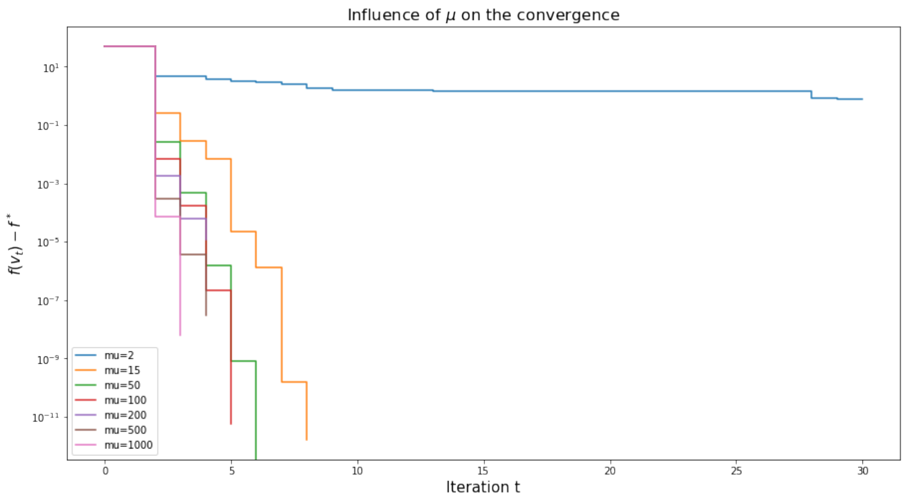

Figure 1: Evolution of the duality gap along Barrier method's iterations for different values of μ. **Parameters:** $\alpha = 0.25$, $\beta = 0.9$, tolerance: $\epsilon = 10^{-7}$. The minimum f^* was attained for $\mu = 50$

values of the backtracking hyperparameters α and β and the same impact of μ on the dual gap was observed.

Assume that strong duality hold. To go from the dual solution ν^* to the primal solution w^*,we use the KKT optimality conditions and then we have to solve this linear system:

$$(X^T X)w^* = X^T(y - \nu^*)$$

To avoid the discussion if $X^T X$ is singular or not,we can use the pseudo-inverse $(X^T X)^+$.
Hence, we can plot, for each μ, the L^1-penalty term $||w^*||_1$
It is obvious in Figure 2 that the norm of w^* "vanishes" to 0 once μ exceeds 2. This is compatible with the sparsity

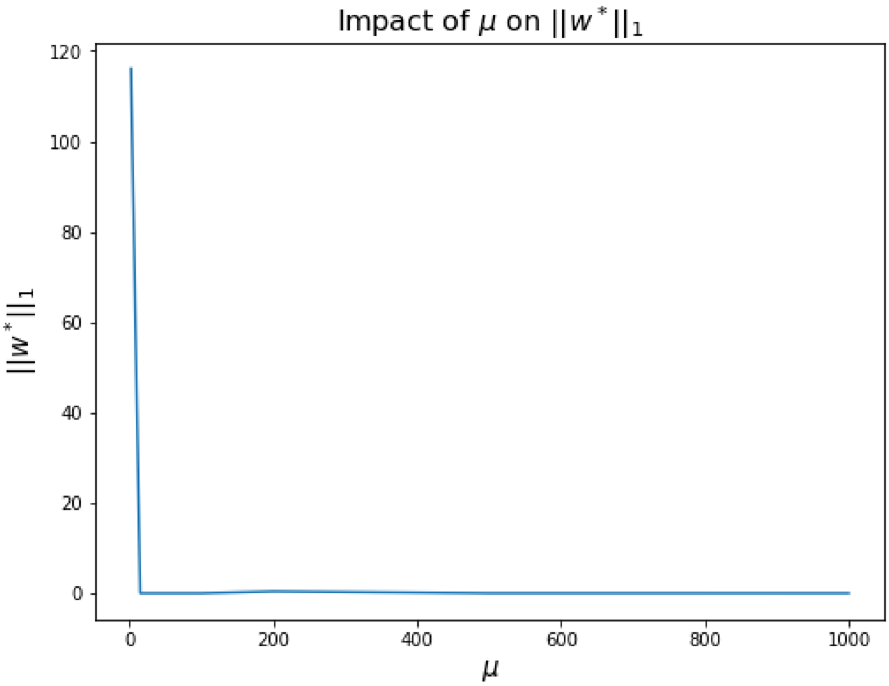

Figure 2: The evolution of $||w^*||_1$ as function of μ. **Parameters:**$\alpha = 0.25$, $\beta = 0.9$, tolerance: $\epsilon = 10^{-7}$.

of w^*. In fact, $(\lambda + sign(w_i^*)(v^{*T}X)_i)|w_i^*| = 0 \; 1 \leq i \leq d$. When μ is high, then, we are more likely to exceed the threshold of having non-zero coefficients for w^*.