Convex Optimization Homework 3

Soël Megdoud

October 2024

1 Exercise 1

1)

Let's introduce a change of variables by setting $z = Xw - y$. This reformulates the original LASSO problem as:

$$\min_{w \in R^d, z \in R^n} \left\{ \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 \right\} \quad \text{subject to } z = Xw - y.$$

The Lagrangian for this constrained problem is:

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

where $\nu \in R^n$ is the dual variable associated with the constraint $z = Xw - y$.

Separating the terms involving w and z, the Lagrangian becomes:

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

To find the dual function $g(\nu)$, we minimize the Lagrangian with respect to w and z separately.

This minimization of the w-dependent part involves the conjugate of the ℓ_1-norm. Using the known conjugate of $\|w\|_1$, we have:

$$\inf_{w \in R^d} \left(\nu^T Xw + \lambda\|w\|_1\right) = \begin{cases} 0, & \text{if } \|X^T\nu\|_\infty \leq \lambda, \\ -\infty, & \text{otherwise.} \end{cases}$$

Thus, the dual variable ν must satisfy $\|X^T\nu\|_\infty \leq \lambda$.

The infimum over z can also bu computed using the conjugate function as well and has been calculated in the previous homework. We have:

$$\inf_{z \in R^n} \left(\frac{1}{2}\|z\|_2^2 - \nu^T z\right) = -\frac{1}{2}\|\nu\|_2^2.$$

Substituting the results of the minimizations into the Lagrangian, the dual problem is to maximize $g(\nu)$, which is equivalent to:

$$\max_{\nu \in R^n} \left\{ -\frac{1}{2}\|\nu\|_2^2 - \nu^T y \right\} \quad \text{subject to } \|X^T\nu\|_\infty \leq \lambda.$$

Turning it to a minimization problem:

$$\min_{\nu \in R^n} \left\{ \frac{1}{2}\|\nu\|_2^2 + \nu^T y \right\} \quad \text{subject to} \quad \|X^T\nu\|_\infty \leq \lambda.$$

The constraint $\|X^T\nu\|_\infty \leq \lambda$ can be rewritten using inequality constraints:

$$\begin{cases} X^T\nu \leq \lambda\mathbf{1}_d, \\ -X^T\nu \leq \lambda\mathbf{1}_d, \end{cases}$$

Now, we can write the dual problem in the standard QP form:

$$\min_{\nu \in R^n} \quad \frac{1}{2}\nu^T\nu + y^T\nu$$
$$\text{subject to} \quad A\nu \leq b,$$

where the constraint matrix A and vector b are:

$$A = \begin{bmatrix} X^T \\ -X^T \end{bmatrix} \in R^{2d\times n}, \quad b = \lambda \begin{bmatrix} \mathbf{1}_d \\ \mathbf{1}_d \end{bmatrix} \in R^{2d}.$$

We can now match the QP standard form:

$$\min_{\nu \in R^n} \quad \nu^T Q \nu + p^T \nu$$
$$\text{subject to} \quad A\nu \leq b,$$

with:

- $Q = \frac{1}{2} I_n$, where I_n is the $n \times n$ identity matrix.
- $p = y$.
- $A = \begin{bmatrix} X^T \\ -X^T \end{bmatrix}$.
- $b = \lambda \begin{bmatrix} \mathbf{1}_d \\ \mathbf{1}_d \end{bmatrix}$.

$Q \succeq 0$ since it is a scaled identity matrix.

2)

Effect of μ on Convergence Precision and Rate.

- Larger values of μ (e.g., $\mu = 50, 100$) result in faster convergence than $\mu = 2$), with fewer iterations needed. However, this seems less stable as for $\mu = 15$ for exemple, the final precision is worse than for $\mu = 2$
- Smaller values of μ (e.g., $\mu = 2$) might result in higher convergence precision.

Effect of μ on w Values.

- The final values of w^* are consistent across all tested values of μ ($\mu = 2, 15, 50, 100$). This consistency arises because the barrier term's relative importance decreases as $t \to \infty$, allowing the optimization to focus primarily on the original quadratic objective function.
- The final solutions w^* are sparse, with many coefficients close to zero, which reflects the expected behavior of LASSO due to the ℓ_1-regularization.

Choice for μ.

For high precision and a more stable convergence, I would choose a small μ like 2.Especially if I have the time to compute all steps. However if time is a constraint I would rather choose a higher μ, it is a trade-off.

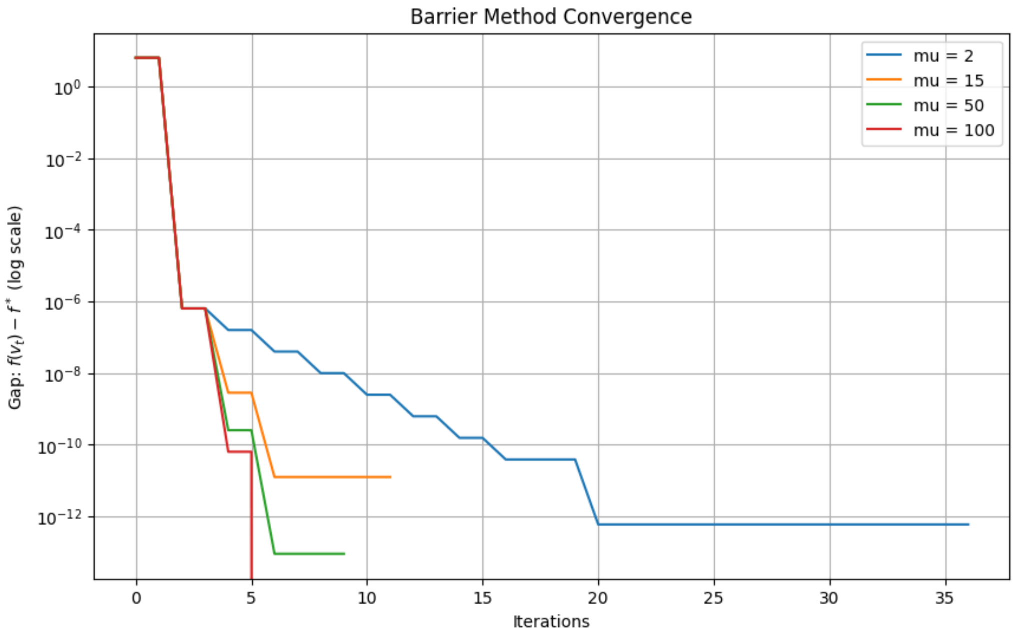

Figure 1: Gap iterations $f(v_t) - f^*$ for Different Values of μ (Log Scale).

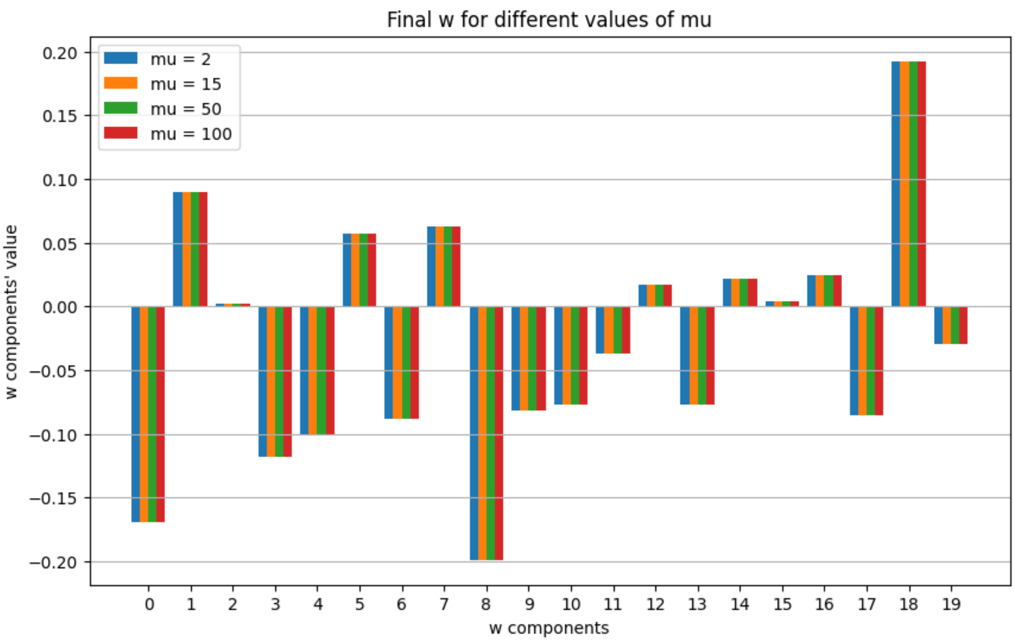

Figure 2: Final values of w^* coefficients for different values of μ.