

Convex Optimization - Homework 3

Samar Rabeh

samar.rabeh@ens-paris-saclay.fr

November , 2024

1 Introduction

The objective of this homework is to solve the LASSO problem. We will approach this by utilizing the QP formulation of the dual LASSO problem. The report is structured as follows: Section 2 outlines the problem setup, while Section 3 presents the theoretical derivation of the QP formulation. In Section 4, we numerically solve the problem using the Barrier method. Lastly, Section 5 discusses and interprets the results obtained.

2 Problem Formulation

We are given $x_1, \dots, x_n \in \mathbb{R}^d$ as data vectors and $y_1, \dots, y_n \in \mathbb{R}$ as the corresponding observations. Our goal is to find regression parameters $w \in \mathbb{R}^d$ that best map the data inputs X to the observations y by minimizing the squared differences. In scenarios where the dimensionality is much higher than the number of observations ($n \ll d$), an ℓ_1 -norm regularization is typically applied to enforce sparsity in the solution. Formally, the LASSO problem can be expressed as:

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} \|Xw - y\|_2^2 + \lambda \|w\|_1$$

where $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{n \times d}$, $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, and $\lambda > 0$ represents the regularization parameter.

3 Dual problem of LASSO

3.1 Dual problem

The LASSO problem can be written as:

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

By introducing the auxiliary variable $z = Xw - y$, we can rewrite this as follows:

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

The Lagrangian for this formulation is :

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

so:

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

The dual function is derived by minimizing the Lagrangian with respect to z and w :

$$g(v) = \inf_{z, w} L(w, z, v)$$

so we can write:

$$g(v) = \inf_z \left(\frac{1}{2} z^T z - v^T z \right) + \inf_w (\lambda \|w\|_1 + v^T Xw) - v^T y$$

For the first part, we minimize:

$$\inf_z \left(\frac{1}{2} z^T z - v^T z \right) = -\frac{1}{2} v^T v$$

For the second -part, we apply the conjugate function of $\|\cdot\|_1$:

$$\inf_w (\lambda \|w\|_1 + v^T Xw) = \begin{cases} 0 & \text{if } \|X^T v\|_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases}$$

Therefore, the dual function is:

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

The dual problem is written as:

$$\max_v \left(-\frac{1}{2} v^T v - v^T y \right) \quad \text{subject to} \quad \|X^T v\|_\infty \leq \lambda$$

3.2 QP formulation

The dual function can be expressed as:

$$g(v) = \inf_{z, w} L(w, z, v)$$

$$g(v) = -\frac{1}{2} v^T v - v^T y \quad \text{if} \quad \|X^T v\|_\infty \leq \lambda$$

Therefore, the dual optimization problem becomes:

$$\max_v \left(-\frac{1}{2}v^T v - v^T y \right) \quad \text{subject to} \quad \|X^T v\|_\infty \leq \lambda$$

This can be reformulated as a quadratic programming (QP) problem:

$$\min_v \frac{1}{2}v^T Q v + p^T v \quad \text{subject to} \quad A v \preceq b$$

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

4 Resolution with Barrier method

4.1 Barrier method applied to our problem

Given strictly feasible v , $t := t(0) > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

Repeat the following steps:

- **Centering step:** Compute $x^*(t)$ by minimizing $f = t f_0 + \varphi$, subject to $Ax = b$.
- **Update:** $x := x^*(t)$.
- **Stopping criterion:** Quit if $\frac{m}{t} < \epsilon$.
- **Increase t :** $t := \mu t$.

4.2 Centering step - Newton's method

Given a starting point $x \in \text{dom} f$, tolerance $\epsilon > 0$.

Repeat:

- Compute the Newton step and decrement.
- Stopping criterion: Quit if $\|\nabla^2 f(x)\|_2^2 \leq \epsilon$.
- Line search: Choose step size t by backtracking line search.
- Update: $x := x + t \Delta x_{nt}$.

5 Experimental results

In our experiments, we simulate data with dimensions $n = 50$ and $d = 100$, representing a high-dimensional scenario. The regularization parameter is set to $\lambda = 10$ to encourage sparsity in the solutions, and we choose a precision level of $\epsilon = 10^{-6}$. The optimization starts from the trivial feasible point $v_0 = 0$. For solving the problem using Newton's method, we implement backtracking line search with fixed parameters $\alpha = 0.01$ and $\beta = 0.5$.

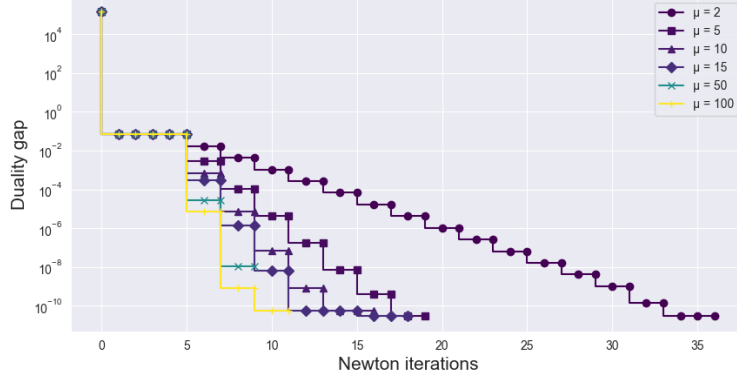


Figure 1: Duality gap over iterations

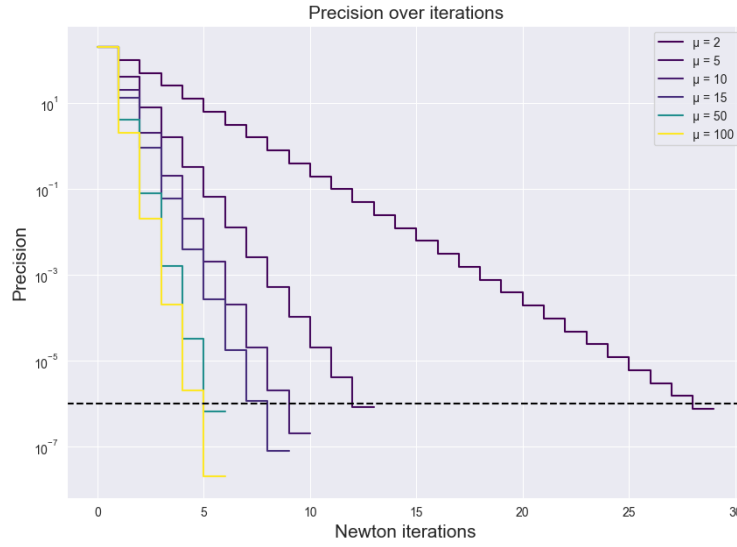


Figure 2: Precision over iterations

5.1 Conclusion

Figure 1 illustrates how the duality gap $f(v_t) - f(v^*)$ evolves as the number of Newton iterations increases, for various values of the barrier method parameter μ in the range of $[2, 5, 10, 15, 30, 50, 100]$.

Figure 2 presents the progression of precision over the cumulative number of Newton steps for the same set of μ values.

From the first plot, it is evident that as μ decreases, the Barrier method converges more slowly, requiring a larger number of iterations to reach convergence (for instance, with $\mu = 2$). This is expected since, with smaller values of μ , each iteration produces smaller updates, leading to a slower reduction in the duality gap. Consequently, more iterations are needed. Therefore, we observe that smaller values of μ correspond to a higher number of outer iterations. On the contrary, for larger values of μ (such as $\mu = 50$ and $\mu = 100$), the method requires fewer outer iterations but longer inner iterations, as the updates are more significant in each step. This results in more iterations for the centering step, which is

carried out using Newton’s method.

In conclusion, there is a trade-off between the number of Barrier method iterations and Newton method iterations. Despite different values of μ , we observe that the algorithm consistently reaches nearly the same final solution. Therefore, to minimize computational cost, it would be advantageous to choose a value of μ around 10 – 15, balancing both outer and inner iterations efficiently.

5.2 Impact of μ on w

Regarding the regression coefficients, we notice that they exhibit significant sparsity, with values falling within the range of $[10^{-13}, 10^{-10}]$ for various choices of μ , as shown in **Figure 3**. These values can be calculated using the formula $w = X^T(v^* + y)$, which follows from the KKT conditions. We do not observe any substantial influence of μ on the coefficients w . Additionally, **Figure 4** illustrates the L_2 norm difference between the computed w and the initial value w_0 , confirming that μ has no significant effect on w . The difference remains in the order of 10^{-8} for all tested values of μ .

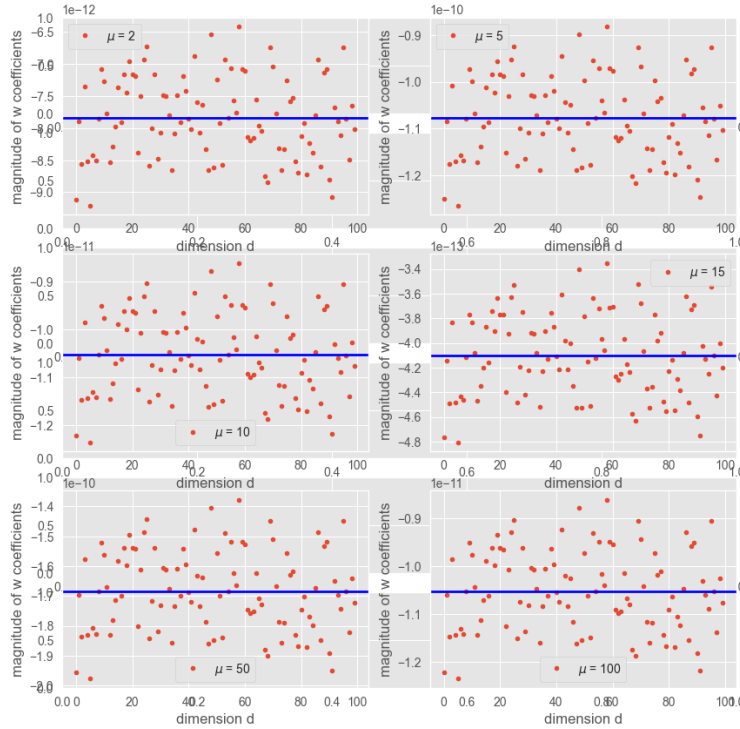


Figure 3: Influence of μ on w magnitude

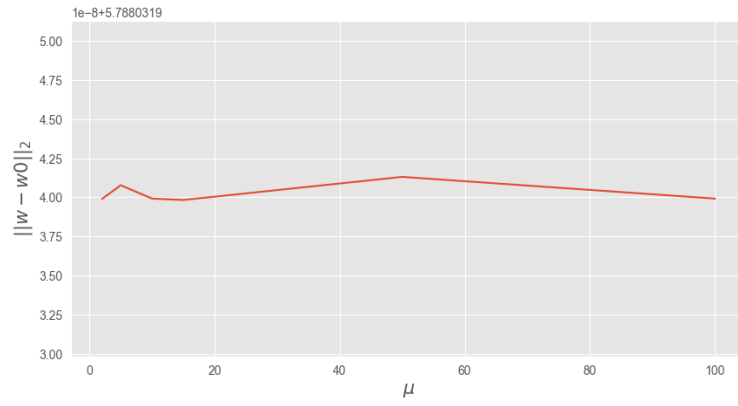


Figure 4: Influence of μ on $\|w - w_0\|_2$