Convex Optimization - Homework 3

DAOUDI Paul

April 21, 2021

We are trying to solve the LASSO problem:

$$\underset{w}{\text{minimize}} \quad \frac{1}{2}\|Xw - y\|_2^2 + \lambda\|w\|_1$$

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

Exercise 1

Deriving the dual

First, let's derive the dual of this problem.

Let's start, as Exercise 2 of Homework 2., by posing $z = Xw - y$ and the problem becomes:

$$\begin{aligned} \underset{z,w}{\text{minimize}} \quad & \frac{1}{2}z^T z + \lambda\|w\|_1 \\ \text{subject to} \quad & z = Xw - y \end{aligned}$$

The Lagrangian is:

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

The dual function $g(v) = \inf_{z,w} \mathcal{L}(z, w, v)$ can be written as:

$$\inf_z \left(z^T z - v^T z\right) + \inf_w \left(\lambda\|w\|_1 + v^T Xw\right) - y^T v$$

The Lagrangian in z is convex so it can be minimized by computing its gradient and setting it to 0. We find $z^\star = v$.

Minimizing with respect to w is slightly more tricky. We can rewrite the elements depending on w as:

$$\inf_w \left(\lambda \|w\|_1 + v^T X w\right) = -\sup_w \left(-w^T \frac{X^T v}{\lambda} - \|w\|_1\right) = -f^\star \left(-\frac{X^T v}{\lambda}\right) = \begin{cases} 0 & \text{if } \|X^T v\|_\infty \leq \lambda; \\ -\infty & \text{else.} \end{cases}$$

where $f^\star$ is the conjugate of the $l1$-norm.

Consequently, the LASSO dual is:

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & -\left(\frac{1}{2} v^T v + y^T v\right) \\ \text{subject to} \quad & \|X^T v\|_\infty \leq \lambda \end{aligned}$$

Now, by setting:

- $Q = \frac{1}{2}\mathbb{I}$,
- $p = y$,
- $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$,
- $b = \lambda_{2d}$

We retrieve the equivalent following Quadratic Problem:

$$\begin{aligned} \underset{v}{\text{minimize}} \quad & v^T Q v + p^T v \\ \text{subject to} \quad & A v \preccurlyeq b \end{aligned} \qquad \text{[QP]}$$

Retrieve the primal solution

Using strong duality of Quadratic Programs, we can retrieve the optimal solution $w^\star$ from its dual $v^\star$. Indeed, since strong duality holds, the KKT conditions must be satisfied at the optimal point $(w^\star, v^\star)$. Especially:

$$\begin{cases} z^\star = v^\star = X w^\star - y & \text{[1.1]} \\ \|w^\star\|_1 - w^\star \left(\frac{X^T v}{\lambda}\right) = 0 & \text{[1.2]} \end{cases}$$

Equation [1.2] guarantees sparsity: it can be written as $\sum_i |w_i^\star| \left(1 - \frac{w_i^\star}{|w_i^\star|} \left(\frac{X^T v}{\lambda}\right)_i\right) = 0$. So, at the optimum, if $\left(\frac{X^T v}{\lambda}\right)_i$ is far from 1 or -1 then $w_i^\star = 0$.

Equation [1.1] gives the remaining coordinates of the primal: $w^\star = \left(X^T X\right)^{-1} X^T (y + v^\star)$.

Note: With the dual, we are now trying to find a n-size vector instead of a d-size vector. Assuming $n << d$, this is a huge gain!!

Exercise 2

Now, let's implement the Barrier method to solve the above Quadratic Program (see Python files).

Small help to understand the Centering Step: In the Barrier method, given t, the Centering Problem is:

$$\underset{v}{\text{minimize}} \quad f(v) = t \cdot \left(v^T Q v + p^T v\right) - \sum_{i=1}^{d} log(b_i - a_i^T v)$$

We can get the analytical formulas of the Gradient and Hessian of f:

$$\nabla f(v) = t\left(Q^T + Q\right) v + tp + A^T z \quad \text{where } z_i = \frac{1}{b_i - a_i^T v}$$
$$H(v) = t\left(Q^T + Q\right) + A^T diag(z)^2 A$$

Note: As requested, we use the Newton's method in the Centering Step. Here is the two important formulas:

$$\Delta v_{nt} = -\nabla^2 f(v)^{-1} \nabla f(v)$$
$$\lambda^2 = \nabla f(v)^T \nabla^2 f(v)^{-1} \nabla f(v)$$