Convex optimization - Homework 3 - LAPASSAT Louis

We consider the following problem (LASSO):

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

where $w \in \mathbb{R}^d$, $X = (x_1, .., x_n) \in \mathbb{R}^{n\times d}$, $y = (y_1, .., y_n) \in \mathbb{R}^n$ and $\lambda > 0$.

Question 1:

We first rewrite (LASSO) problem like so:

$$\begin{aligned} \min_{w,z} \quad & \frac{1}{2}||Xw - y||_2^2 + \lambda||z||_1 \\ \text{s.t.} \quad & z = w \end{aligned}$$

Hence the Lagrangian is:

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

where $\nu \in \mathbb{R}^d$. So the dual function is:

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

As the first part is independent of z, and the second part is independent of w, the minimizers can be found by minimizing both part separately, over w and z respectively. Starting with z we have:

$$\inf_z \lambda||z||_1 - \nu^T z \iff -\sup_z \nu^T z - \lambda||z||_1 = \begin{cases} 0 & \text{if } ||\nu||_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases}.$$

Now for w we have:

$$\inf_w \frac{1}{2}||Xw - y||_2^2 + \nu^T w.$$

Since the function is convex, we can just consider derivative with respect to w in order to get the infimum.

$$\nabla_w L(w, z, \nu) = X^T(Xw - y) + \nu \text{ and } \nabla_w^2 L(x, z, \nu) = X^T X.$$

Since $X^T X \geq 0$ we know that a stationary point is a global minima. Therefore w^* solves:

$$\nabla_w L(w, z, \nu)|_{w^*} = 0 \iff X^T X w^* = X^T y - \nu,$$

and if X has independent columns: $w^* = (X^T X)^{-1}(X^T y - \nu)$. Finally the dual problem is given by:

$$\max_\nu g(\nu) \iff \begin{aligned} \max_\nu \quad & \frac{1}{2}||Xw^* - y||_2^2 + \nu^T w^* \\ \text{s.t.} \quad & ||\nu||_\infty \leq \lambda \end{aligned} \qquad (D)$$

Now we want to simplify and transform (D) into a general Quadratic Problem (QP). To do so let's first replace w^*, then drop all terms that do not depend on ν (the $\arg\max$ remain the same):

$$\frac{1}{2}||Xw^* - y||_2^2 + \nu^T w^* \implies -\frac{1}{2}\nu^T(X^T X)^{-1}\nu + \nu^T(X^T X)^{-1}X^T y.$$

So (D) is equivalent to the following minimization problem:

$$\begin{array}{ll} \min\limits_{\nu} & \frac{1}{2}\nu^T(X^TX)^{-1}\nu - \nu^T(X^TX)^{-1}X^Ty \\ \text{s.t.} & ||\nu||_\infty \leq \lambda \end{array} .$$

Now we only have to transform the constraints:

$$||\nu||_\infty \leq \lambda \Longleftrightarrow -\lambda \leq \nu_i \leq \lambda \quad \forall i = 1,..,d \Longleftrightarrow \begin{bmatrix} I_d \\ -I_d \end{bmatrix} \nu \leq \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

where I_d is the $d \times d$ identity matrix. Finally the problem can be rewritten like so:

$$\begin{array}{ll} \min\limits_{v} & v^TQv + p^Tv \\ \text{s.t.} & Av \leq b \end{array} \quad (QP),$$

where $Q = \frac{1}{2}(X^TX)^{-1}$, $p = -(X^TX)^{-1}X^Ty$, $A = \begin{bmatrix} I_d \\ -I_d \end{bmatrix}$ and $b = [\lambda, .., \lambda]^T \in \mathbb{R}^{2d}$.