

Also observe that for an optimal solution, we have

$$\begin{aligned} \frac{1}{2} \|y - Xw\|_2^2 + w^\top X^\top (y - Xw) &= \frac{1}{2} \|y\|_2^2 - w^\top X^\top y + \frac{1}{2} w^\top X^\top Xw + w^\top X^\top y - w^\top X^\top Xw \\ &= \frac{1}{2} (\|y\|_2^2 - \|Xw\|_2^2) \\ &= \frac{1}{2} (\|y\|_2^2 - \|y - \lambda\|_2^2) = G(\lambda). \end{aligned}$$

Since the objective function is convex and the constraints are qualified, by Theorem 50.19(2) the duality gap is zero, so for optimal solutions of the primal and the dual,  $G(\lambda) = L(\xi, w, \epsilon)$ , that is

$$\frac{1}{2} \|y - Xw\|_2^2 + w^\top X^\top (y - Xw) = \frac{1}{2} \|\xi\|_2^2 + \tau \|w\|_1 = \frac{1}{2} \|y - Xw\|_2^2 + \tau \|w\|_1,$$

which yields the equation

$$w^\top X^\top (y - Xw) = \tau \|w\|_1. \quad (**_1)$$

The above is the inner product of  $w$  and  $X^\top (y - Xw)$ , so whenever  $w_i \neq 0$ , since  $\|w\|_1 = |w_1| + \cdots + |w_n|$ , in view of (\*), we must have  $(X^\top (y - Xw))_i = \tau \operatorname{sgn}(w_i)$ . If

$$S = \{i \in \{1, \dots, n\} \mid w_i \neq 0\}, \quad (\dagger)$$

if  $X_S$  denotes the matrix consisting of the columns of  $X$  indexed by  $S$ , and if  $w_S$  denotes the vector consisting of the nonzero components of  $w$ , then we have

$$X_S^\top (y - X_S w_S) = \tau \operatorname{sgn}(w_S). \quad (**_2)$$

We also have

$$\|X_{\bar{S}}^\top (y - X_S w_S)\|_\infty \leq \tau, \quad (**_3)$$

where  $\bar{S}$  is the complement of  $S$ .

Equation (\*\*<sub>2</sub>) yields

$$X_S^\top X_S w_S = X_S^\top y - \tau \operatorname{sgn}(w_S),$$

so if  $X_S^\top X_S$  is invertible (which will be the case if the columns of  $X$  are linearly independent), we get

$$w_S = (X_S^\top X_S)^{-1} (X_S^\top y - \tau \operatorname{sgn}(w_S)). \quad (**_4)$$

In theory, if we know the support of  $w$  and the signs of its components, then  $w_S$  is determined, but in practice this is useless since the problem is to find the support and the sign of the solution.