Since $\alpha_+, \alpha_- \geq 0$, for any $i \in \{1, \dots, n\}$ the minimum of $(\alpha_+)_i - (\alpha_-)_i$ is $-\tau$, and the maximum is τ. If we recall that for any $z \in \mathbb{R}^n$,

$$\|z\|_\infty = \max_{1 \leq i \leq n} |z_i|,$$

it follows that the constraints

$$\begin{aligned} \alpha_+ + \alpha_- &= \tau \mathbf{1}_n \\ X^\top \lambda &= \alpha_+ - \alpha_- \end{aligned}$$

are equivalent to

$$\left\|X^\top \lambda\right\|_\infty \leq \tau.$$

The above is equivalent to the $2n$ constraints

$$-\tau \leq (X^\top \lambda)_i \leq \tau, \quad 1 \leq i \leq n.$$

Therefore, the dual lasso program is given by

$$\begin{aligned} &\text{maximize} \quad -\frac{1}{2}\left(\|y - \lambda\|_2^2 - \|y\|_2^2\right) \\ &\text{subject to} \\ &\qquad\qquad \left\|X^\top \lambda\right\|_\infty \leq \tau, \end{aligned}$$

which (since $\|y\|_2^2$ is a constant term) is equivalent to

Program (Dlasso2):

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|y - \lambda\|_2^2 \\ &\text{subject to} \\ &\qquad\qquad \left\|X^\top \lambda\right\|_\infty \leq \tau, \end{aligned}$$

minimizing over $\lambda \in \mathbb{R}^m$.

One way to solve lasso regression is to use the dual program to find $\lambda = \xi$, and then to use linear programming to find w by solving the linear program arising from the lasso primal by holding ξ constant. The best way is to use ADMM as explained in Section 52.8(4). There are also a number of variations of gradient descent; see Hastie, Tibshirani, and Wainwright [89].

In theory, if we know the support of w and the signs of its components, then w is determined as we now explain.

In view of the constraint $y - Xw = \xi$ and the fact that for an optimal solution we must have $\xi = \lambda$, the following condition must hold:

$$\left\|X^\top (Xw - y)\right\|_\infty \leq \tau. \tag{$*$}$$