so by setting the gradient $\nabla L_{\xi,w,\epsilon,b}$ to zero we obtain the equations

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

Thus b is also determined, and the dual lasso program is identical to the first lasso dual (**Dlasso2**), namely

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

minimizing over λ.

Since the equations $\xi = \lambda$ and

$$y - Xw - b\mathbf{1}_m = \xi$$

hold, from $Cb = \mathbf{1}_m^\top \lambda$ we get

$$\frac{1}{m}\mathbf{1}_m^\top y - \frac{1}{m}\mathbf{1}_m^\top Xw - b\frac{1}{m}\mathbf{1}_m^\top \mathbf{1} = \frac{1}{m}\mathbf{1}_m^\top \lambda,$$

that is

$$\overline{y} - (\overline{X^1} \ \cdots \ \overline{X^n})w - b = \frac{C}{m}b,$$

which yields

$$b = \frac{m}{m+C}(\overline{y} - (\overline{X^1} \ \cdots \ \overline{X^n})w).$$

As in the case of ridge regression, a defect of the approach where b is also penalized is that the solution for b is not invariant under adding a constant c to each value y_i

It is interesting to compare the behavior of the methods:

1. Ridge regression (**RR6′**) (which is equivalent to (**RR3**)).
2. Ridge regression (**RR3**b), with b penalized (by adding the term Kb^2 to the objective function).
3. Least squares applied to $[X \ \ \mathbf{1}]$.
4. (**lasso3**).