with $y, \xi, \mathbf{1} \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$. Note that in Program (**RR3**) minimization is performed over ξ, w and b, but b is not penalized in the objective function. As in Section 55.1, the objective function is strictly convex.

The Lagrangian associated with this program is

$$L(\xi, w, b, \lambda) = \xi^\top \xi + K w^\top w - w^\top X^\top \lambda - \xi^\top \lambda - b\mathbf{1}^\top \lambda + \lambda^\top y.$$

Since L is (strictly) convex as a function of ξ, b, w, by Theorem 40.13(4), it has a minimum iff $\nabla L_{\xi,b,w} = 0$. We get

$$\begin{aligned}\lambda &= 2\xi\\ \mathbf{1}^\top \lambda &= 0\\ w &= \frac{1}{2K} X^\top \lambda = X^\top \frac{\xi}{K}.\end{aligned}$$

As before, if we set $\xi = K\alpha$ we obtain $\lambda = 2K\alpha$, $w = X^\top \alpha$, and

$$G(\alpha) = -K\alpha^\top (XX^\top + KI_m)\alpha + 2K\alpha^\top y.$$

Since $K > 0$ and $\lambda = 2K\alpha$, the dual to ridge regression is the following program

Program (DRR3):

$$\begin{aligned}&\text{minimize} \quad \alpha^\top (XX^\top + KI_m)\alpha - 2\alpha^\top y\\ &\text{subject to}\\ &\qquad\qquad \mathbf{1}^\top \alpha = 0,\end{aligned}$$

where the minimization is over α.

Observe that up to the factor 1/2, this problem satisfies the conditions of Proposition 42.3 with

$$\begin{aligned}A &= (XX^\top + KI_m)^{-1}\\ b &= y\\ B &= \mathbf{1}_m\\ f &= 0,\end{aligned}$$

and x renamed as α. Therefore, it has a unique solution (α, μ) (beware that $\lambda = 2K\alpha$ is **not** the λ used in Proposition 42.3, which we rename as μ), which is the unique solution of the KKT-equations

$$\begin{pmatrix} XX^\top + KI_m & \mathbf{1}_m \\ \mathbf{1}_m^\top & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \mu \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

Since the solution given by Proposition 42.3 is

$$\mu = (B^\top AB)^{-1}(B^\top Ab - f), \quad \alpha = A(b - B\mu),$$