function is strictly convex so if an optimal solution exists, then it is unique; the proof is left as an exercise.

The Lagrangian associated with this optimization problem is

$$\begin{aligned}L(\xi, w, \epsilon, b, \lambda, \alpha_+, \alpha_-) &= \frac{1}{2}\xi^\top \xi - \xi^\top \lambda + \lambda^\top y - b\mathbf{1}_m^\top \lambda \\ &\quad + \epsilon^\top(\tau \mathbf{1}_n - \alpha_+ - \alpha_-) + w^\top(\alpha_+ - \alpha_- - X^\top \lambda) + \frac{1}{2}K w^\top w,\end{aligned}$$

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

$$\begin{aligned}\xi &= \lambda \\ Kw &= -(\alpha_+ - \alpha_- - X^\top \lambda) \qquad (*_w) \\ \alpha_+ + \alpha_- - \tau \mathbf{1}_n &= 0 \\ \mathbf{1}_m^\top \lambda &= 0.\end{aligned}$$

We find that $(*_w)$ determines w. Using these equations, we can find the dual function but in order to solve the dual using ADMM, since $\lambda \in \mathbb{R}^m$, it is more convenient to write $\lambda = \lambda_+ - \lambda_-$, with $\lambda_+, \lambda_- \in \mathbb{R}^m_+$ (recall that $\alpha_+, \alpha_- \in \mathbb{R}^n_+$). As in the derivation of the dual of ridge regression, we rescale our variables by defining $\beta_+, \beta_-, \mu_+, \mu_-$ such that

$$\alpha_+ = K\beta_+, \;\; \alpha_- = K\beta_-, \;\; \lambda_+ = K\mu_+, \;\; \lambda_- = K\mu_-.$$

We also let $\mu = \mu_+ - \mu_-$ so that $\lambda = K\mu$. Then $\mathbf{1}_m^\top \lambda = 0$ is equivalent to

$$\mathbf{1}_m^\top \mu_+ - \mathbf{1}_m^\top \mu_- = 0,$$

and since $\xi = \lambda = K\mu$, we have

$$\begin{aligned}\xi &= K(\mu_+ - \mu_-) \\ \beta_+ + \beta_- &= \frac{\tau}{K}\mathbf{1}_n.\end{aligned}$$

Using $(*_w)$ we can write

$$\begin{aligned}w &= -(\beta_+ - \beta_- - X^\top \mu) = -\beta_+ + \beta_- + X^\top \mu_+ - X^\top \mu_- \\ &= \begin{pmatrix} -I_n & I_n & X^\top & -X^\top \end{pmatrix} \begin{pmatrix} \beta_+ \\ \beta_- \\ \mu_+ \\ \mu_- \end{pmatrix}.\end{aligned}$$