we compute $\nabla L_{w,\epsilon,\xi,b,\eta}$. The gradient $\nabla L_{w,\epsilon,\xi,b,\eta}$ is given by

$$\nabla L_{w,\epsilon,\xi,b,\eta} = \begin{pmatrix} w + X\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \\ 2K_s\epsilon - \lambda \\ 2K_s\xi - \mu \\ b + \mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu \\ \mathbf{1}_p^\top \lambda + \mathbf{1}_q^\top \mu - \nu \end{pmatrix}.$$

By setting $\nabla L_{w,\epsilon,\xi,b,\eta} = 0$ we get the equations

$$\begin{aligned} w &= -X\begin{pmatrix}\lambda \\ \mu\end{pmatrix}, & (*_w) \\ 2K_s\epsilon &= \lambda \\ 2K_s\xi &= \mu \\ b &= -(\mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu) \\ \mathbf{1}_p^\top \lambda + \mathbf{1}_q^\top \mu &= \nu. \end{aligned}$$

As we said earlier, both w an b are determined by λ and μ. We can use the equations to obtain the following expression for the dual function $G(\lambda, \mu, \gamma)$,

$$\begin{aligned} G(\lambda,\mu,\gamma) &= -\frac{1}{4K_s}(\lambda^\top\lambda + \mu^\top\mu) - \frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix} X^\top X \begin{pmatrix}\lambda \\ \mu\end{pmatrix} - \frac{b^2}{2} \\ &= -\frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix}\left(X^\top X + \begin{pmatrix}\mathbf{1}_p\mathbf{1}_p^\top & -\mathbf{1}_p\mathbf{1}_q^\top \\ -\mathbf{1}_q\mathbf{1}_p^\top & \mathbf{1}_q\mathbf{1}_q^\top\end{pmatrix} + \frac{1}{2K_s} I_{p+q}\right)\begin{pmatrix}\lambda \\ \mu\end{pmatrix}. \end{aligned}$$

Consequently the dual program is equivalent to the minimization program

Dual of the Soft margin SVM (SVM$_{s5}$):

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\begin{pmatrix}\lambda^\top & \mu^\top\end{pmatrix}\left(X^\top X + \begin{pmatrix}\mathbf{1}_p\mathbf{1}_p^\top & -\mathbf{1}_p\mathbf{1}_q^\top \\ -\mathbf{1}_q\mathbf{1}_p^\top & \mathbf{1}_q\mathbf{1}_q^\top\end{pmatrix} + \frac{1}{2K_s} I_{p+q}\right)\begin{pmatrix}\lambda \\ \mu\end{pmatrix} \\ &\text{subject to} \\ &\qquad \sum_{i=1}^p \lambda_i + \sum_{j=1}^q \mu_j = \nu \\ &\qquad \lambda_i \geq 0, \quad i = 1, \ldots, p \\ &\qquad \mu_j \geq 0, \quad j = 1, \ldots, q. \end{aligned}$$

It is shown in Section 54.16 how the dual program is solved using ADMM from Section 52.6. If the primal problem is solvable, this yields solutions for λ and μ.

The constraint

$$\sum_{i=1}^p \lambda_i + \sum_{j=1}^q \mu_j = \nu$$