Step 5: Write the dual program in matrix form.

Maximizing the dual function $G(\lambda, \mu)$ over its domain of definition is equivalent to maximizing

$$\widehat{G}(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}$$

subject to the constraint

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

so we formulate the dual program as,

$$\begin{aligned} & \text{maximize} \quad -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} \\ & \text{subject to} \\ & \qquad \sum_{i=1}^{p} \lambda_i = \sum_{j=1}^{q} \mu_j \\ & \qquad \lambda \geq 0, \ \mu \geq 0, \end{aligned}$$

or equivalently,

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} - \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} \\ & \text{subject to} \\ & \qquad \sum_{i=1}^{p} \lambda_i = \sum_{j=1}^{q} \mu_j \\ & \qquad \lambda \geq 0, \ \mu \geq 0. \end{aligned}$$

The constraints of the dual program are a lot simpler than the constraints

$$\begin{pmatrix} X^\top & \begin{matrix} \mathbf{1}_p \\ -\mathbf{1}_q \end{matrix} \end{pmatrix} \begin{pmatrix} w \\ b \end{pmatrix} \leq -\mathbf{1}_{p+q}$$

of the primal program because these constraints have been "absorbed" by the objective function $\widehat{G}(\lambda, \nu)$ of the dual program which involves the matrix $X^\top X$. The matrix $X^\top X$ is symmetric positive semidefinite, but not invertible in general.

Step 6: Solve the dual program.

This step involves using numerical procedures typically based on gradient descent to find λ and μ, for example, ADMM from Section 52.6. Once λ and μ are determined, w is determined by $(*_1)$ and b is determined as in Section 50.6 using the fact that there is at least some i_0 such that $\lambda_{i_0} > 0$ and some j_0 such that $\mu_{j_0} > 0$.

Remarks: