and open, we can apply Theorem 40.13, which gives a necessary and sufficient condition for a minimum. The gradient of $L(w, b, \lambda, \mu)$ with respect to w and b is

$$\begin{aligned}\nabla L_{w,b} &= \begin{pmatrix} I_n & 0_n \\ 0_n^\top & 0 \end{pmatrix} \begin{pmatrix} w \\ b \end{pmatrix} + \begin{pmatrix} X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \\ \mathbf{1}_p^\top \lambda \quad -\mathbf{1}_q^\top \mu \end{pmatrix} \\ &= \begin{pmatrix} w \\ 0 \end{pmatrix} + \begin{pmatrix} X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \\ \mathbf{1}_p^\top \lambda \quad -\mathbf{1}_q^\top \mu \end{pmatrix}.\end{aligned}$$

The necessary and sufficient condition for a minimum is

$$\nabla L_{w,b} = 0,$$

which yields

$$w = -X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \tag{$*_1$}$$

and

$$\mathbf{1}_p^\top \lambda - \mathbf{1}_q^\top \mu = 0. \tag{$*_2$}$$

The second equation can be written as

$$\sum_{i=1}^p \lambda_i = \sum_{j=1}^q \mu_j. \tag{$*_3$}$$

Plugging back w from $(*_1)$ into the Lagrangian and using $(*_2)$ we get

$$G(\lambda, \mu) = -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q}; \tag{$*_4$}$$

of course, $\begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} = \sum_{i=1}^p \lambda_i + \sum_{j=1}^q \mu_j$. Actually, to be perfectly rigorous $G(\lambda, \mu)$ is only defined on the intersection of the hyperplane of equation $\sum_{i=1}^p \lambda_i = \sum_{j=1}^q \mu_j$ with the convex octant in $\mathbb{R}^{p+q}$ given by $\lambda \geq 0, \mu \geq 0$, so for all $\lambda \in \mathbb{R}_+^p$ and all $\mu \in \mathbb{R}_+^q$, we have

$$G(\lambda, \mu) = \begin{cases} -\frac{1}{2} \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} X^\top X \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \lambda^\top & \mu^\top \end{pmatrix} \mathbf{1}_{p+q} & \text{if } \sum_{i=1}^p \lambda_i = \sum_{j=1}^q \mu_j \\ -\infty & \text{otherwise.} \end{cases}$$

Note that the condition

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

is Condition $(*_2)$ of Example 50.6, which is not surprising.