Solutions to Part A of Problem Sheet 9

Solution (9.1) The problem is

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + \mu \sum_{i=1}^n s_i \\
&\text{subject to} \quad y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - 1 + s_i \geq 0, \quad 1 \leq i \leq n \\
&\quad s_i \geq 0.
\end{aligned}$$

Writing $\boldsymbol{X}$ for the matrix with $y_i \boldsymbol{x}_i$ as rows, $\boldsymbol{y}$ and $\boldsymbol{s}$ for the vectors with the y_i and s_i as entries, and $\boldsymbol{e}$ for the vector with all ones, we get the compact form

$$\begin{aligned}
&\text{minimize} \quad \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + \mu \boldsymbol{s}^\top \boldsymbol{e} \\
&\text{subject to} \quad -\boldsymbol{X}\boldsymbol{w} - b\boldsymbol{y} + \boldsymbol{e} - \boldsymbol{s} \leq \boldsymbol{0} \\
&\quad -\boldsymbol{s} \leq \boldsymbol{0},
\end{aligned}$$

where we also changed the signs to get "$\leq$" inequalities. The Lagrangian of the problem involves the primal variables $\boldsymbol{w}, b, \boldsymbol{s}$, and *two* sets of dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$, corresponding to the two sets of inequality constraints:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{s}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}) &= \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + \mu \boldsymbol{s}^\top \boldsymbol{e} - \boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{w} - b\boldsymbol{\lambda}^\top \boldsymbol{y} + \boldsymbol{\lambda}^\top(\boldsymbol{e} - \boldsymbol{s}) - \tilde{\boldsymbol{\lambda}}^\top \boldsymbol{s} \\
&= \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} - \boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{w} - b\boldsymbol{\lambda}^\top \boldsymbol{y} + \boldsymbol{s}^\top(\mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}) + \boldsymbol{\lambda}^\top \boldsymbol{e}.
\end{aligned}$$

The second of these expression collects all the terms that involve each of the variables $\boldsymbol{w}$, $\boldsymbol{s}$ and b. When trying to minimize this expression with respect to $\boldsymbol{w}, \boldsymbol{s}, b$, we notice that the minimum is $-\infty$ if $\boldsymbol{\lambda}^\top \boldsymbol{y} \neq 0$ or $\mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}} \neq \boldsymbol{0}$. For values of $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ such that $\mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}} = \boldsymbol{0}$ and $\boldsymbol{\lambda}^\top \boldsymbol{y} = 0$, we determine the minimum by computing the gradient and setting it to zero:

$$\begin{aligned}
\nabla_{\boldsymbol{x}} \mathcal{L} &= \boldsymbol{w} - \boldsymbol{X}^\top \boldsymbol{\lambda} = \boldsymbol{0} \\
\nabla_{\boldsymbol{s}} \mathcal{L} &= \mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}} = \boldsymbol{0} \\
\frac{\partial \mathcal{L}}{\partial b} &= \boldsymbol{y}^\top \boldsymbol{\lambda} = 0
\end{aligned}$$

As in the lecture, we can replace $\boldsymbol{w}$ in the Lagrangian by $\boldsymbol{X}^\top \boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^\top \boldsymbol{y}$ by 0. In addition, we get to replace $\mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}$ by $\boldsymbol{0}$, overall giving the following expression for $\boldsymbol{g}$:

$$\boldsymbol{g}(\boldsymbol{\lambda}) = \begin{cases} -\frac{1}{2}\boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top \boldsymbol{e} & \text{if } \mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^\top \boldsymbol{y} = 0 \\ -\infty & \text{else.} \end{cases}$$

It turns out that the objective of the Lagrange dual is the same as in the setting without the $\boldsymbol{s}$ variables. A difference becomes apparent when considering the Lagrange dual problem, which requires maximizing over $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\tilde{\boldsymbol{\lambda}} \geq \mathbf{0}$. Since $\mu \boldsymbol{e} - \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}} \geq \mathbf{0}$, we conclude that $\boldsymbol{\lambda} \leq \mu \boldsymbol{e}$, i.e., every λ_i is bounded by μ. We can therefore rephrase the Lagrange dual problem as

$$\text{minimize} \quad \frac{1}{2}\boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \boldsymbol{e} \qquad \text{subject to} \quad \mathbf{0} \leq \boldsymbol{\lambda} \leq \mu \boldsymbol{e}.$$

It is interesting to note what happens when μ approach 0 or ∞, how this affects the behaviour of the primal and of the Lagrange dual problem!

The KKT conditions are found by collecting the constraints of the primal problem, of the dual problem, the gradient of the Lagrangian, and complementarity slackness:

$$\begin{aligned}
\boldsymbol{X}\boldsymbol{w} + b\boldsymbol{y} - \boldsymbol{e} + \boldsymbol{s} &\geq \mathbf{0} \\
\boldsymbol{s} &\geq \mathbf{0} \\
\boldsymbol{\lambda} &\geq \mathbf{0} \\
\tilde{\boldsymbol{\lambda}} &\geq \mathbf{0} \\
\lambda_i(1 - y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b)) &= 0 \text{ for } 1 \leq i \leq n \\
\boldsymbol{w} - \boldsymbol{X}^\top \boldsymbol{\lambda} &= \mathbf{0} \\
\nabla_{\boldsymbol{s}} \mathcal{L} &= \mu \boldsymbol{e} - \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}} = \mathbf{0} \\
\boldsymbol{y}^\top \boldsymbol{\lambda} &= 0.
\end{aligned}$$

A closer look at these conditions reveals that some of the variables (for example, $\tilde{\boldsymbol{\lambda}}$) could be eliminated, thus simplifying the system.

Solution (9.2) Write

$$\boldsymbol{P}(\lambda) = \boldsymbol{P} + \sum_{i=1}^m \lambda_i \boldsymbol{P}_i, \ \boldsymbol{q}(\lambda) = \boldsymbol{q} + \sum_{i=1}^m \lambda_i \boldsymbol{q}_i, \ \boldsymbol{r}(\lambda) = \boldsymbol{r} + \sum_{i=1}^m \lambda_i \boldsymbol{r}_i.$$

With this notation, we can express the Lagrangian as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}(\lambda)\boldsymbol{x} + \boldsymbol{q}(\lambda)^\top \boldsymbol{x} + \boldsymbol{r}(\lambda).$$

We can now approach this minimization problem just as we would approach any such problem with a positive semidefinite matrix: compute the gradient in $\boldsymbol{x}$, $\boldsymbol{P}(\boldsymbol{\lambda}) + \boldsymbol{q}(\boldsymbol{\lambda})$, and set this to zero. Plugging in the result, $\boldsymbol{x} = -\boldsymbol{P}(\boldsymbol{\lambda})^{-1}\boldsymbol{q}(\boldsymbol{\lambda})$, into the equation for the Lagrangian, we get for $\boldsymbol{\lambda} \geq \mathbf{0}$

$$g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{q}(\boldsymbol{\lambda})^\top \boldsymbol{P}(\boldsymbol{\lambda})^{-1}\boldsymbol{q}(\boldsymbol{\lambda}) + \boldsymbol{r}(\boldsymbol{\lambda}).$$

The Lagrange dual is then given by

$$\begin{aligned}
&\text{maximize} \quad -\frac{1}{2}\boldsymbol{q}(\boldsymbol{\lambda})^\top \boldsymbol{P}(\boldsymbol{\lambda})^{-1}\boldsymbol{q}(\boldsymbol{\lambda}) + \boldsymbol{r}(\boldsymbol{\lambda}) \\
&\text{subject to} \quad \boldsymbol{\lambda} \geq \mathbf{0}.
\end{aligned}$$

This function looks simpler at first sight, since it only involves non-negativity constraints, but it requires the inverse of a linear combination of the matrices $\boldsymbol{P}_i$, which makes things less straight-forward.

Solution (9.3) First of all, we can make the objective function linear by writing the problem as

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} + r - t \leq 0 \\ & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^\top \boldsymbol{x} + r_i \leq 0, \ 1 \leq i \leq m. \end{aligned}$$

Now we can factor each $\boldsymbol{P}_i$ as $\boldsymbol{P}_i = \boldsymbol{M}_i^\top \boldsymbol{M}_i$. The matrix

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{M}_i\boldsymbol{x} \\ \boldsymbol{x}^\top \boldsymbol{M}_i & -r_i - \boldsymbol{q}_i^\top \boldsymbol{x} \end{pmatrix}$$

is positive semidefinite if and only if $\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^\top \boldsymbol{x} + r_i \leq 0$, by the hint. We can therefore write the QCQP as

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \begin{pmatrix} \boldsymbol{I} & \boldsymbol{M}\boldsymbol{x} \\ \boldsymbol{x}^\top \boldsymbol{M} & -r - \boldsymbol{q}^\top \boldsymbol{x} + t \end{pmatrix} \succeq \boldsymbol{0} \\ & \begin{pmatrix} \boldsymbol{I} & \boldsymbol{M}_i\boldsymbol{x} \\ \boldsymbol{x}^\top \boldsymbol{M}_i & -r_i - \boldsymbol{q}_i^\top \boldsymbol{x} \end{pmatrix} \succeq \boldsymbol{0}, \ 1 \leq i \leq m. \end{aligned}$$

We can formulate several semidefinite constraints as one by assembling the above matrices as diagonal blocks of a big matrix.