# Solutions to Part A of Problem Sheet 8

**Solution (8.1)** We write the optimization problem slightly different as

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) + \frac{1}{t}\varphi(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq \mathbf{0} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

Using the fact that the gradient of $\varphi(\boldsymbol{x})$ is given by

$$\nabla_{\boldsymbol{x}}\varphi(\boldsymbol{x}) = -\sum_{i=1}^{m} \frac{\nabla f_i(\boldsymbol{x})}{f_i(\boldsymbol{x})},$$

we can derive the KKT conditions for this problem as

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}) &\leq \mathbf{0} \\ \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b} \\ \tilde{\boldsymbol{\lambda}} &\geq \mathbf{0} \\ \tilde{\lambda}_i f_i(\boldsymbol{x}) &= 0,\ 1 \leq i \leq m \\ \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*) - \frac{1}{t}\sum_{i=1}^{m} \frac{1}{f_i(\boldsymbol{x})}\nabla f_i(\boldsymbol{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\mu} &= \mathbf{0}, \end{aligned}$$

We now join the coefficients of $\nabla f_i(\boldsymbol{x})$ in the last line and set

$$\lambda_i := \tilde{\lambda}_i - \frac{1}{t f_i(\boldsymbol{x})}.$$

The last equation then becomes

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}) + \sum_{i=1}^{m} + \boldsymbol{A}^\top \boldsymbol{\mu} = \mathbf{0}.$$

These new multipliers also satisfy

$$\lambda_i f_i(\boldsymbol{x}) = -\frac{1}{t}.$$

Rewriting the KKT conditions for the barrier problem in terms of the $\lambda_i$ gives the modified KKT conditions.

**Solution (8.2)** For this problem we just have to figure out how to modify the derivations done in class to the current problem, essentially adding the extra term involving the $s_i$ in the objective, replacing the $-1$ with $-1 + s_i$ in the constraints, and adding some non-negativity constraints for the $s_i$. The Lagrangian of the problem is then

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

This might look slightly involved, but breaking it down we can recognise all the relevant terms in there as coming from either the objective function or the constraints.

To find the Lagrange dual $g(\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}})$, we set the gradient to 0,

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

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 $\boldsymbol{g}$:

$$\boldsymbol{g}(\boldsymbol{\lambda}) = -\frac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top \boldsymbol{e}.$$

It turns out that the Lagrange dual is the same! A difference becomes apparent when considering the Lagrange dual problem, which requires maximizing over $\boldsymbol{\lambda} \geq \boldsymbol{0}$ and $\tilde{\boldsymbol{\lambda}} \geq \boldsymbol{0}$. Since $\mu \boldsymbol{e} - \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}} \geq \boldsymbol{0}$, we conclude that $\boldsymbol{\lambda} \leq \mu \boldsymbol{e}$, i.e., every $\lambda_i$ is bounded by $\mu$. We can therefore rephrase the Lagrange dual problem as

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

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:

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

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 (8.3)** 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.