Solutions to Part A of Problem Sheet 6

Solution (6.1) The claim is that the neighbourhood

$$\mathcal{N}_{-\infty}(1) = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \in \mathcal{F}^\circ : x_i s_i \geq \mu\}$$

coincides with the central path $\mathcal{C}$. Clearly, since μ is the *average* of the $x_i s_i$, $x_i s_i \geq \mu$ for all i must imply $x_i s_i = \mu$ for all i (we can't all be better or equal than average, unless we are all equal). But then, such a vector is clearly on the central path. Conversely, if $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \in \mathcal{C}$, then there exists a $\tau > 0$ such that $x_i s_i = \tau$ for all i. But then, $\mu = \frac{1}{n}\sum_{i=1}^n x_i s_i = \frac{1}{n}\sum_{i=1}^n \tau = \tau = x_i s_i$ for all i, so that $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \in \mathcal{N}_{-\infty}(1)$.

Solution (6.2) If $f(\boldsymbol{x})$ is convex and $\boldsymbol{x}$ and $\boldsymbol{y}$ are such that $f(\boldsymbol{x}) \leq 0$ and $f(\boldsymbol{y}) \leq 0$, then

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq (1-\lambda) f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) \leq 0,$$

so that the set is convex. If we denote by $\mathcal{C}_i = \{\boldsymbol{x} : g_i(\boldsymbol{x}) \leq 0\}$ and $\mathcal{D}_j = \{\boldsymbol{x} : h_j(\boldsymbol{x}) = 0\}$, then

$$\mathcal{C} = \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_m \cap \mathcal{D}_1 \cap \cdots \cap \mathcal{D}_\ell$$

is an intersection of convex sets, and therefore convex.

Solution (6.3) The Lagrangian of the quadratic problem is given by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} - \sum_{i=1}^m \lambda_i(\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle - b_i).$$

The gradient of the Lagrangian is

$$\boldsymbol{Q}\boldsymbol{x} - \sum_{i=1}^m \lambda_i \boldsymbol{a}_i = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{0},$$

where we denoted by $\boldsymbol{a}_i$ the columns of the matrix $\boldsymbol{A}^\top$ (so that $\boldsymbol{a}_i^\top$ are the rows of $\boldsymbol{A}$). Assuming that $\boldsymbol{Q}$ is invertible, we get the equation for $\boldsymbol{x}$

$$\boldsymbol{x} = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\lambda}. \tag{1}$$

This would be a closed form solution for $\boldsymbol{x}$, were it not for the yet unknown Lagrange multipliers $\boldsymbol{\lambda}$. We can, however, get an expression for the Lagrange multipliers in terms of the known data. For this, we multiply (1) with $\boldsymbol{A}$ to get

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\lambda},$$

which holds at an optimal point (since the constraints $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ are expected to hold). Note that the only unknown parameter in this equation is the vector of Lagrange multipliers $\boldsymbol{\lambda}$, all the rest depends on the known quantities $\boldsymbol{b}$, $\boldsymbol{Q}$, and $\boldsymbol{A}$. Solving this $m \times m$ system of linear equations for λ we get

$$\boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b},$$

and plugging this into (1), we get the closed form solution for $\boldsymbol{x}$ as

$$\boldsymbol{x} = \boldsymbol{Q}^{-1}\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\top})^{-1}\boldsymbol{b}.$$

In practice, computing $\boldsymbol{x}$ this way may not be very efficient due to conditioning and computational complexity issues, and one would solve the resulting system of equations that gives $\boldsymbol{\lambda}$ using some matrix factorizations.