Solutions to Part A of Problem Sheet 7

Solution (7.1) 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 (7.2) 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.

Solution (7.3) The problem is not convex since the equality constraint is not linear, and the inequality constraint is not convex. We can formulate an equivalent convex optimization problem as

$$\text{minimize} \quad x_1^2 + x_2^2 \qquad \text{subject to} \quad x_1 \leq 0,\ x_1 + x_2 = 0.$$

Solution (7.4) First of all, note that the function is only defined for $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$. This is the *domain* of the function.

We introduce new variables $\boldsymbol{y}$ and derive the dual to the problem

$$\begin{aligned} \text{minimize} \quad & -\sum_{i=1}^{m} \log(y_i) \\ \text{subject to} \quad & \boldsymbol{y} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

Note that by restricting to the domain of the problem, we don't have to explicitly ask for $\boldsymbol{y}$ to be non-negative: the objective function wouldn't make sense for negative values.

The Lagrangian to this problem is

$$\begin{aligned} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\mu}) &= -\sum_{i=1}^{m} \log(y_i) + \boldsymbol{\mu}^\top(\boldsymbol{y} - \boldsymbol{b} + \boldsymbol{A}\boldsymbol{x}) \\ &= \sum_{i=1}^{m} -\log(y_i) + \mu_i(y_i - b_i + \boldsymbol{a}_i^\top \boldsymbol{x}). \end{aligned}$$

The dual function is

$$g(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}, \boldsymbol{y}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\mu}),$$

where the infimum is taken over the domain of $\mathcal{L}$ (in particular, this requires $\boldsymbol{y} \geq \boldsymbol{0}$). The infimum is $-\infty$ if $\boldsymbol{\mu}^\top \boldsymbol{A} \neq \boldsymbol{0}$. If $\boldsymbol{\mu}$ has negative terms, then the infimum is also $-\infty$ (we could then choose an arbitrary large value for the corresponding y variable).

If $\boldsymbol{\mu} > 0$, then we can determine the minimum by computing the gradient. For the partial derivative in y_i we get

$$\frac{\partial \mathcal{L}}{\partial y_i} = -\frac{1}{y_i} + \mu_i = 0,$$

so at the minimum we have $y_i = \frac{1}{\mu_i}$. For the gradient in the $\boldsymbol{x}$ variables we get $\nabla_{\boldsymbol{x}} \mathcal{L} = \boldsymbol{A}^\top \boldsymbol{\mu} = \boldsymbol{0}$. It follows that the dual function is

$$g(\boldsymbol{\mu}) = \begin{cases} \sum_{i=1}^{m} \log(\mu_i) + m - \boldsymbol{b}^\top \boldsymbol{\mu} & \text{if } \boldsymbol{A}^\top \boldsymbol{\mu} = \boldsymbol{0},\ \boldsymbol{\mu} > 0, \\ -\infty & \text{else,} \end{cases}$$

where we used that $\log(y_i) = \log(1/\mu_i) = -\log(\mu_i)$.