Chalmers/GU
Mathematics

EXAM SOLUTION

TMA947/MAN280
APPLIED OPTIMIZATION

Date: 07–03–12
Examiner: Michael Patriksson

Question 1

(LP duality)

(2p) a) The dual version of the problem is given by:

$$
\begin{array}{lrl}
\text{maximize} & w = 3y_1 + 4y_2, & \\
\text{subject to} & y_1 - 2y_2 & \leq 2, \\
& 2y_1 + y_2 & \leq 1, \\
& 3y_1 & \leq \alpha, \\
& -y_1 + 3y_2 & \leq 0, \\
& 3y_1 - 2y_2 & \leq 1, \\
& y_1, \quad y_2 & \geq 0.
\end{array}
$$

With $\alpha = 1$, sketching the feasible set (see Figure 1) shows that $\boldsymbol{y}^* = (\frac{1}{3}, \frac{1}{9})^{\mathrm{T}}$ with $w^* = \frac{13}{9}$. From the graph it is easy to se that the constraints corresponding to x_1, x_2 and x_5 are not active, hence $x_1^* = x_2^* = x_5^* = 0$ must hold. Both dual variables are stricly positive in the optimum point, and therefore both primal constraints must hold with equality. For this to happen, we must have $x_3^* = \frac{13}{9}$ and $x_4^* = \frac{4}{3}$. This gives (as expected) $z^* = \frac{13}{9} = w^*$.

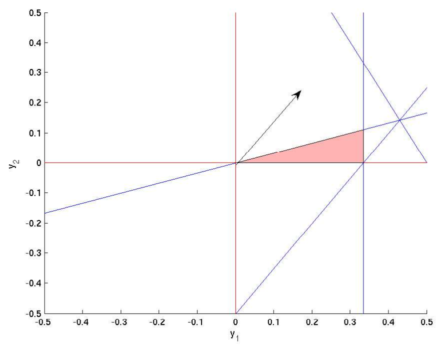

Figure 1: The dual feasible set.

(1p) b) Changing α means moving the vertical constraint horizontally. We see that if $\alpha < 0$ the dual feasible set is empty and $\boldsymbol{x}^*$ can no longer be optimal due to strong duality. If we choose $\alpha > \frac{9}{7}$ the vertical constraint becomes inactive and x_3^* can no longer stay positive due to complementarity. For $\alpha \in [0, \frac{9}{7}]$ the dual basis remain optimal and so does $\boldsymbol{x}^*$.

(3p) Question 2

(sufficiency of the KKT conditions under convexity)

See Theorem 5.45.

(3p) Question 3

(Farkas Lemma)

Show that $\{\, \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{0}^m, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} < 0 \,\}$ is inconsistent by showing that $\{\, \boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}^n, \boldsymbol{y} \geq \boldsymbol{0}^m \,\}$ has a solution $\boldsymbol{y} = (1,1)^{\mathrm{T}}$.

(3p) Question 4

(the Frank–Wolfe algorithm)

Starting at $\boldsymbol{x}_0 = (1,2)^{\mathrm{T}}$, the algorithm proceeds as follows: $f(\boldsymbol{x}_0) = 41$; $\nabla f(\boldsymbol{x}_0) = (40,2)^{\mathrm{T}}$; $\boldsymbol{y}_0 = (0,0)^{\mathrm{T}}$; the lower bound $z(\boldsymbol{y}_0) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^{\mathrm{T}}(\boldsymbol{y}_0 - \boldsymbol{x}_0) = -3$; $\boldsymbol{p}_0 = \boldsymbol{y}_0 - \boldsymbol{x}_0 = -(1,2)^{\mathrm{T}}$; $f(\boldsymbol{x}_0 + \alpha\boldsymbol{p}_0) = 10(2-\alpha)^2 + (1-2\alpha)^2$, which yields minimum $\alpha = 1$ over the interval $\alpha \in [0,1]$; $\boldsymbol{x}_1 = (0,0)^{\mathrm{T}}$; $f(\boldsymbol{x}_1) = 11$; $\nabla f(\boldsymbol{x}_1) = (20,-2)^{\mathrm{T}}$; $\boldsymbol{y}_1 = (0,2)^{\mathrm{T}}$; the lower bound is $z(\boldsymbol{y}_1) = f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^{\mathrm{T}}(\boldsymbol{y}_1 - \boldsymbol{x}_1) = 7$; $\boldsymbol{p}_1 = (0,2)^{\mathrm{T}}$; $f(\boldsymbol{x}_1 + \alpha\boldsymbol{p}_1) = 10 + (2\alpha - 1)^2$, which yields minimum in $\alpha = 0.5$; $\boldsymbol{x}_2 = (0,1)^{\mathrm{T}}$; $f(\boldsymbol{x}_2) = 10$; $\nabla f(\boldsymbol{x}_2) = (20,0)^{\mathrm{T}}$; $\boldsymbol{y}_2 = (0,0)^{\mathrm{T}}$ (for example). The lower and upper bounds are equal, hence $\boldsymbol{x}_2 = (0,1)^{\mathrm{T}} = \boldsymbol{x}^*$, with optimal value $f^* = 10$.

Question 5

(the Levenberg–Marquardt modification of Newton's method)

(1p) a) The Slater CQ holds, so the optimum must be a KKT point. Use the KKT conditions to get

$$\begin{aligned}
&\nabla^2 f(\boldsymbol{x})\boldsymbol{p} + \nabla f(\boldsymbol{x}) + 2\mu\boldsymbol{p} = (\nabla^2 f(\boldsymbol{x}) + 2\mu \boldsymbol{I}^n)\boldsymbol{p} + \nabla f(\boldsymbol{x}) = \boldsymbol{0}^n,\\
&\mu\,(\|\boldsymbol{p}\|^2 - \delta) = 0,\\
&\mu \geq 0.
\end{aligned}$$

(1p) b) The Hessian at $\boldsymbol{x}^0$ is

$$\nabla^2 f(\boldsymbol{x}^0) = \begin{bmatrix} -4 & 0 \\ 0 & 2 \end{bmatrix},$$

which requires a shift $\sigma > 4$, otherwise the step $\boldsymbol{p}$ is not a descent direction.

(1p) c) Set $\boldsymbol{p} = -\alpha \nabla f(\boldsymbol{x})$, and compute the optimal value of α:

$$\begin{aligned}
\text{minimize} \quad & \phi(\alpha) = \frac{1}{2}\alpha^2 \nabla f(\boldsymbol{x})\nabla^2 f(\boldsymbol{x})\nabla f(\boldsymbol{x}) + \alpha\|\nabla f(\boldsymbol{x})\|^2\\
\text{subject to} \quad & \alpha \leq \frac{\sqrt{\delta}}{\|\nabla f(\boldsymbol{x})\|}
\end{aligned}$$

Which gives

$$\alpha = \text{minimum}\left\{\frac{\sqrt{\delta}}{\|\nabla f(\boldsymbol{x})\|}, \frac{\|\nabla f(\boldsymbol{x})\|^2}{\nabla f(\boldsymbol{x})^{\mathrm{T}}\nabla^2 f(\boldsymbol{x})\,\nabla f(\boldsymbol{x})}\right\}.$$

(3p) Question 6

(modelling)

Introduce the variables:

x_j	Amount of coal in tonnes to be transp to mill j, $j = 1, 2$
y_{ij}	Amount of ore in tonnes to be transp from mine i to mill j, $i = 1, 2, 3; j = 1, 2$
u_j	Amount of energy in kWh used in mill j, $j = 1, 2$
v_j	Produced amount of steel in mill j, $j = 1, 2$
w_{jk}	Number of produced units of product k in mill j, $j = 1, 2; k = 1, 2$ (where $k = 1$ represents plates and $k = 2$ represents pipes)

With the notation in the problem formulation, the objective is

$$\min \quad \sum_j (g + r_j)x_j + \sum_i \sum_j (h + t_{ij})y_{ij} + \sum_j pu_j + \sum_j qv_j - \sum_j \sum_k s_k w_{jk}$$

and the constraints are

$$\sum_j y_{ij} \leq cp_i, \qquad i = 1, 2, 3, \tag{1}$$

$$\sum_j w_{jk} \leq d_k, \qquad k = 1, 2, \tag{2}$$

$$v_j \leq \frac{x_j}{a}, \qquad j = 1, 2, \tag{3}$$

$$v_j \leq \sum_i \frac{y_{ij}}{b}, \qquad j = 1, 2, \tag{4}$$

$$v_j \leq \frac{u_j}{c}, \qquad j = 1, 2, \tag{5}$$

$$v_j \geq \sum_k \frac{w_{jk}}{e_k}, \qquad j = 1, 2, \tag{6}$$

$$x_j, y_{ij}, u_j, v_j, w_{jk} \geq 0, \qquad \forall i, j, k, \tag{7}$$

where (1) is the capacity constraint in the ore mines, (2) is the limitation of the market demand, (3)–(5) are the process constrint telling how much raw material that at least is needed, (6) the balance constraint in the production of product using the steel and finally (7) the logical non-negativity constraints on all variables.

Question 7

(linear programming duality and optimality)

(1p) a) Let the Lagrange multipliers be denoted by $\boldsymbol{\mu} \in \mathbb{R}^m_+$ and $\boldsymbol{\sigma} \in \mathbb{R}^n_+$, respectively.

Setting the partial derivative of the Lagrangian $L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\sigma}) := \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) - \boldsymbol{\sigma}^{\mathrm{T}}\boldsymbol{x}$ to zero yields that $\boldsymbol{\sigma} = \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu}$ must hold. (This can be used to eliminate $\boldsymbol{\sigma}$ altogether.) Inserting this into the Lagrangian function yields that the optimal value of the Lagrangian when minimized over $\boldsymbol{x} \in \mathbb{R}^n$ is $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}$. According to the construction of the Lagrangian dual problem, $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}$ should then be maximized over the constraints that the dual variables are non-negative; here, we obtain that $\boldsymbol{\mu} \geq \boldsymbol{0}^m$, and from $\boldsymbol{\sigma} \geq \boldsymbol{0}^n$ we further obtain that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} \leq \boldsymbol{c}$ must hold. The Lagrangian dual problem hence is equivalent to the canonical LP dual:

$$\begin{aligned} \text{maximize} \quad & w = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}, \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} \leq \boldsymbol{c}, \\ & \boldsymbol{\mu} \geq 0^n. \end{aligned} \tag{D}$$

(2p) b) We identify $X = \mathbb{R}^n$, $\ell = m + n$, and the vector

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \\ -\boldsymbol{x} \end{pmatrix}.$$

The optimality conditions of (1) include both multiplier vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, but $\boldsymbol{\sigma}$ is eliminated here as well. Primal feasibility corresponds to the requirements that $\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$ and $\boldsymbol{x} \geq \boldsymbol{0}^n$ hold, while dual feasibility was above shown to be equivalent to the requirements that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} \leq \boldsymbol{c}$ and $\boldsymbol{\mu} \geq \boldsymbol{0}^m$ hold. Finally, complementarity yields that $\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = 0$ hold, as well as the condition that $\boldsymbol{\sigma}^{\mathrm{T}}\boldsymbol{x} = 0$ holds; the latter reduces (thanks to the possibility to eliminate $\boldsymbol{\sigma}$) to $\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{c}) = 0$, the familiar one. We are done.