Chalmers/Gothenburg University
Mathematical Sciences

EXAM SOLUTION

TMA947/MMG621
OPTIMIZATION, BASIC COURSE

Date: 12–12–17
Examiner: Michael Patriksson

Question 1

(the simplex method)

(2p) a) We first rewrite the problem on standard form. We rewrite $x_2 = x_2^+ - x_2^-$ and introduce slack variables s_1 and s_2.

$$\begin{aligned} f^* = \text{infimum} \quad & -2x_1 + x_2^+ - x_2^-, \\ \text{subject to} \quad & -x_1 + x_2^+ - x_2^- + s_1 = 0, \\ & x_1 - 2x_2^+ + 2x_2^- + s_2 = 4, \\ & x_1, x_2^+, x_2^-, s_1, s_2 \geq 0. \end{aligned}$$

Phase I

If we start with basis (s_1, s_2), we have a unit matrix.

Phase II

Calculating the reduced costs, we obtain $\tilde{\boldsymbol{c}}_N = (-2, 1, -1)^{\mathrm{T}}$, meaning that x_1 should enter the basis. From the minimum ratio test, we get that the only eligible outgoing variable is s_2.

Updating the basis we now have (x_1, s_1) in the basis. At this BFS, we have that $\tilde{\boldsymbol{c}}_N = (-3, 3, 2)^{\mathrm{T}}$, meaning that x_2^+ should enter the basis. Performing the minimum ratio test, we see that $\boldsymbol{B}^{-1}\boldsymbol{N}_{x_2^+} = (-2, -1)^{\mathrm{T}}$, which means that the problem is unbounded. A direction of unboundedness in variables in the standard form then is

$$\boldsymbol{p} = \begin{bmatrix} -\boldsymbol{B}^{-1}\boldsymbol{N}_{x_2^+} \\ \boldsymbol{e}_{x_2^+} \end{bmatrix} = \begin{bmatrix} p_{x_1} \\ p_{s_1} \\ p_{x_2^+} \\ p_{x_2^-} \\ p_{s_2} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Translating this to the original variables, we see that a direction of unboundedness is $\boldsymbol{p} = \begin{bmatrix} p_{x_1} \\ p_{x_2} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

(1p) b) We have that $f^* = -\infty$, since the problem is unbounded. By weak duality, we have that the LP dual is infeasible. But the feasible set to the dual problem does not depend on the right-hand side vector $\boldsymbol{b}$, so the dual problem will always be infeasible. The only thing that can affect f^* is if the perturbation also makes the primal problem infeasible, meaning that $f^* = \infty$. However, in this example, the problem will always be unbounded.

Question 2

(Lagrangian duality)

(2p) a) We create the Lagrangian function

$$L(\boldsymbol{x}, \mu) = (x_1-1)^2-2x_2+\mu(2x_2-x_1-2) = \left(x_1^2 - 2x_1 - \mu x_1\right)+(2(\mu-1)x_2)+1-2\mu.$$

The dual function then is

$$q(\mu) = \min_{\boldsymbol{x}\geq 0} L(\boldsymbol{x}, \mu) = 1 - 2\mu + \min_{x_1\geq 0}\left(x_1^2 - 2x_1 - \mu x_1\right) + \min_{x_2\geq 0}\left(2(\mu-1)x_2\right).$$

At $\mu = 0$, since the objective function coefficent for x_2 is negative, letting $x_2 \to \infty$ yields unbounded solutions to the Lagrangian subproblem. Thus $q(0) = -\infty$.

At $\mu = 2$, to minimize the convex quadratic problem in x_1 we let $x_1 = 1 + \mu/2 = 2$, and $x_2 = 0$. Thus $q(2) = -7$. By weak duality it follows that $q(2) \leq f^*$.

To find an upper bound, choose any feasible point, e.g. $(x_1, x_2) = (1, 1)$, which has objective value -2. Hence $f^* \in [-7, -2]$.

(1p) b) Take $\boldsymbol{g}^1, \boldsymbol{g}^2 \in \partial f(\boldsymbol{x})$ and $\lambda \in (0, 1)$. Then

$$\begin{aligned}
f(\boldsymbol{x}) + \left(\lambda\boldsymbol{g}^1 + (1-\lambda)\boldsymbol{g}^2\right)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) &= f(\boldsymbol{x}) + \lambda(\boldsymbol{g}^1)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) + (1-\lambda)(\boldsymbol{g}^2)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \\
&= \lambda \underbrace{\left[f(\boldsymbol{x}) + (\boldsymbol{g}^1)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x})\right]}_{\leq f(\boldsymbol{y})} + (1-\lambda) \underbrace{\left[f(\boldsymbol{x}) + (\boldsymbol{g}^2)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x})\right]}_{\leq f(\boldsymbol{y})} \\
&\leq \lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{y}) = f(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^n.
\end{aligned}$$

So $\lambda\boldsymbol{g}^1 + (1-\lambda)\boldsymbol{g}^2 \in \partial f(\boldsymbol{x})$, which implies that $\partial f(\boldsymbol{x})$ is a convex set.

(3p) Question 3

(gradient projection)

The starting point is $\boldsymbol{x}_0 = (0\ 2)^{\mathrm{T}}$, where $f(\boldsymbol{x}_0) = 8$. At this point, $\nabla f(\boldsymbol{x}_0) = (2\ 2)^{\mathrm{T}}$, so the search direction is $\boldsymbol{p}_0 = (-2\ \ -2)^{\mathrm{T}}$. With the step length $\alpha = \frac{1}{4}$, we obtain the point $\boldsymbol{x} = (-\frac{1}{2}\ \frac{3}{2})^{\mathrm{T}}$; as it is infeasible, we need to project this point onto the feasible set; this yield the new iteration point $\boldsymbol{x}_1 = (0\ \frac{3}{2})^{\mathrm{T}}$. The objective value at $\boldsymbol{x}_1$ is 9/2, so in this instance the step length was short enough to produce descent.

We are then asked to check whether $\boldsymbol{x}_1$ is a stationary point, or indeed an optimal solution. As the gradient projection method is a descent method, we simply generate the search direction from $\boldsymbol{x}_1$ to find out if descent is obtained or not.

At $\boldsymbol{x}_1$, we have that $\nabla f(\boldsymbol{x}_1) = (-3\ 6)^{\mathrm{T}}$, so the next search direction hence is $\boldsymbol{p}_0 = (3\ \ -6)^{\mathrm{T}}$. At $\boldsymbol{x}_1$ this is feasible descent direction. Hence, $\boldsymbol{x}_1$ cannot be optimal.

Question 4

(KKT conditions)

(1p) a) Let $f(\boldsymbol{x}^*) = x_1 + x_2$, $g_1(\boldsymbol{x}) = -x_1$, $g_2(\boldsymbol{x}) = -x_2$, $g_3(\boldsymbol{x}) = x_1x_2$. We get that $\nabla g_1(\boldsymbol{x}^*) = [-1, 0]^{\mathrm{T}}$, $\nabla g_2(\boldsymbol{x}^*) = [0, -1]^{\mathrm{T}}$, $\nabla g_3(\boldsymbol{x}^*) = [0, 0]^{\mathrm{T}}$ and $\nabla f(\boldsymbol{x}^*) = [1, 1]^{\mathrm{T}}$. Hence

$$\nabla f(\boldsymbol{x}^*) + 1\nabla g_1(\boldsymbol{x}^*) + 1\nabla g_2(\boldsymbol{x}^*) = 0,$$

which shows that $\boldsymbol{x}^*$ is a KKT point.

(1p) b) The only (locally) optimal solution is $\boldsymbol{x}^* = (0, 0)^{\mathrm{T}}$. The feasible set S consists of the non-negative coordinates axes, and hence for any $\boldsymbol{x} \in S$ it holds that for $\boldsymbol{p} = \boldsymbol{x} - \boldsymbol{x}^*$ we have $p_1p_2 = 0$. Thus $p_1p_2 = 0$ for any $\boldsymbol{p} \in T_S(\boldsymbol{x}^*)$. Since both $[1, 0]^{\mathrm{T}} \in T_S(\boldsymbol{x}^*)$ and $[0, 1] \in T_S(\boldsymbol{x}^*)$ but $(1/2, 1/2) \notin T_S(\boldsymbol{x}^*)$ it follows that $T_S(\boldsymbol{x}^*)$ is *not* a convex set. On the other hand, $G(\boldsymbol{x}^*)$ is a convex polyhedron. Hence $T_S(\boldsymbol{x}^*) \neq G(\boldsymbol{x}^*)$.

(1p) c) Let $\boldsymbol{x}^*$ be locally optimal. Then the geometric optimality condition yields that $\mathring{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset$. For any $\boldsymbol{p} \in \mathrm{conv} T_S(\boldsymbol{x}^*)$ we have $\boldsymbol{p} = \sum_{j=1}^k \alpha_j \boldsymbol{p}_j$ for some $\boldsymbol{p}_j \in T_S(\boldsymbol{x}^*)$, $0 \leq \alpha_j$, $j = 1, \dots, k$, $\sum_{j=1}^k \alpha_j = 1$. Then

$$\begin{aligned}\nabla f(x^*)^{\mathrm{T}} \boldsymbol{p} &= \nabla f(x^*)^{\mathrm{T}} \sum_{j=1}^k \alpha_j \boldsymbol{p}_j \\ &= \sum_{j=1}^k \underbrace{\alpha_j}_{\geq 0} \underbrace{\nabla f(x^*)^{\mathrm{T}} \boldsymbol{p}_j}_{\geq 0} \geq 0,\end{aligned}$$

since $\boldsymbol{p}_j \in T_S(\boldsymbol{x}^*) \implies \boldsymbol{p}_j \notin \mathring{F}(\boldsymbol{x}^*)$. Hence $\boldsymbol{p} \notin \mathring{F}(\boldsymbol{x}^*)$, and thus $\mathring{F}(\boldsymbol{x}^*) \cap \text{conv}T_S(\boldsymbol{x}^*) = \emptyset$. By the Guignard CQ, then $\mathring{F}(\boldsymbol{x}^*) \cap G(\boldsymbol{x}^*) = \emptyset$. The rest of the proof follows by Farkas' Lemma as for the proof under Abadies CQ.

Finally we note that for the problem in (a), $G(\boldsymbol{x}^*) = \{\boldsymbol{p} \mid p_1, p_2 \geq 0\} = \{\boldsymbol{p} \mid \boldsymbol{p} = \alpha[1,0]^{\mathrm{T}} + \beta[0,1]^{\mathrm{T}}, \alpha, \beta \geq 0\}$. But since $[0,1]^{\mathrm{T}} \in T_S(\boldsymbol{x}^*)$ and $[1,0]^{\mathrm{T}} \in T_S(\boldsymbol{x}^*)$ it follows that $G(\boldsymbol{x}^*) \subseteq \text{convT}_{\text{S}}(\boldsymbol{x}^*)$. But since $\text{convT}_{\text{S}}(\boldsymbol{x}^*) \subseteq \text{G}(\boldsymbol{x}^*)$ always holds, we must have that $G(\boldsymbol{x}^*) = \text{convT}_{\text{S}}(\boldsymbol{x}^*)$; the Guignard CQ holds at $\boldsymbol{x}^*$. At any other feasible $\boldsymbol{x}$, it is easy to see that the Abadie CQ holds, hence the Guignard CQ holds everywhere.

Question 5

(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.

(3p) Question 6

(convergence of an exterior penalty method)

This is Theorem 13.3.

(3p) Question 7

Introduce the binary variables x_{ij} for $i, j \in \mathcal{N}$, denoting the value placed in row i column j in the solution to the puzzle. Further introduce the variables

$$y_{i_1 i_2 j} = \begin{cases} 1, & \text{rows } i_1 \text{ and } i_2 \text{ are identical in column } j, \\ 0, & \text{otherwise.} \end{cases}$$

and

$$z_{j_1 j_2 i} = \begin{cases} 1, & \text{columns } j_1 \text{ and } j_2 \text{ are identical in row } i, \\ 0, & \text{otherwise.} \end{cases}$$

for $i_1, i_2, j_1, j_2 \in \mathcal{N}$, $i_1 < i_2$, $j_1 < j_2$. A puzzle solution is equivalent to a feasible solution to the constraints

$$\sum_{j=k}^{k+2} x_{ij} \geq 1, \qquad i \in \mathcal{N}, k = 1, \dots, n-2, \qquad (1)$$

$$\sum_{j=k}^{k+2} x_{ij} \leq 2, \qquad i \in \mathcal{N}, k = 1, \dots, n-2, \qquad (2)$$

$$\sum_{j=1}^{n} x_{ij} = \frac{n}{2}, \qquad i \in \mathcal{N}, \qquad (3)$$

$$\sum_{i=k}^{k+2} x_{ij} \geq 1, \qquad j \in \mathcal{N}, k = 1, \dots, n-2, \qquad (4)$$

$$
\begin{aligned}
\sum_{i=k}^{k+2} x_{ij} &\leq 2, && j \in \mathcal{N}, k = 1, \dots, n-2, && (5)\\
\sum_{i=1}^{n} x_{ij} &= \frac{n}{2}, && j \in \mathcal{N}, && (6)\\
y_{i_1 i_2 j} &\geq x_{i_1 j} + x_{i_2 j} - 1, && i_1, i_2, j \in \mathcal{N}, i_1 < i_2, && (7)\\
y_{i_1 i_2 j} &\geq 1 - (x_{i_1 j} + x_{i_2 j}), && i_1, i_2, j \in \mathcal{N}, i_1 < i_2, && (8)\\
z_{j_1 j_2 i} &\geq x_{i j_1} + x_{i j_2} - 1, && j_1, j_2, j \in \mathcal{N}, j_1 < j_2, && (9)\\
z_{j_1 j_2 i} &\geq 1 - (x_{i j_1} + x_{i j_2}), && j_1, j_2, j \in \mathcal{N}, j_1 < j_2, && (10)\\
\sum_{j=1}^{n} y_{i_1 i_2 j} &\leq n - 1, && i_1, i_2 \in \mathcal{N}, i_1 < i_2, && (11)\\
\sum_{i=1}^{n} z_{j_1 j_2 i} &\leq n - 1, && j_1, j_2 \in \mathcal{N}, j_1 < j_2, && (12)\\
x_{ij} &= a_{ij}, && (i,j) \in \mathcal{D}, && (13)\\
x_{ij}, y_{i_1 i_2 j}, z_{j_1 j_2 i} &\in \{0,1\}, && i, j \in \mathcal{N}, j_1, j_2 \in \mathcal{N}, j_1 < j_2. && (14)
\end{aligned}
$$

The first three constraints correspond, in order, to requiring that in a fixed row, three consecutive numbers cannot all be 0, three consecutive numbers cannot not all be 1, and half the numbers in the row must be 1. Constraints (4)–(6) state the similar logic over individual columns. Constraints (7)–(8) enforce that $y_{i_1 i_2 j} = 1$ if and only if $x_{i_1 j} = x_{i_2 j}$; the right hand side of (7) evaluates to 1 if $x_{i_1 j} = x_{i_2 j} = 1$, and 0 otherwise, the right hand side of (8) evaluates to 1 if $x_{i_1 j} = x_{i_2 j} = 0$, and 0 otherwise. Constraints (9)–(10) state the similar logic as (7)–(8) over columns instead of rows. Finally, constraints (11)–(12) state that two rows/columns cannot be identical everywhere, while (13) ensures that we respect the inital puzzle data.

To verify uniqueness of solutions, first solve the model with an arbitrary (linear) objective function. Denote (if it exists) the optimal puzzle solution by $\bar{x}_{ij}$, $i, j \in \mathcal{N}$. Now consider the objective function to minimize $f(\boldsymbol{x}, \mathbf{y}, \mathbf{z}) = \sum_{i,j \in \mathcal{N}} c_{ij} x_{ij}$, where $c_{ij} = \bar{x}_{ij}$. Resolve the model with this objective function. If the puzzle solution is unique, the optimal value $f^* = n^2/2$ (the number of ones in the puzzle solution $\bar{x}_{ij}$). If the puzzle solution in not unique, there exists a solution that places a 0 at some point where $\bar{x}_{ij} = 1$, and hence $f^* < n^2/2$.