Chalmers/GU
Mathematics

EXAM

TMA947/MAN280
OPTIMIZATION, BASIC COURSE

Date:	07–03–12
Time:	House V, morning
Aids:	Text memory-less calculator, English–Swedish dictionary
Number of questions:	7; passed on one question requires 2 points of 3. Questions are *not* numbered by difficulty. To pass requires 10 points and three passed questions.
Examiner:	Michael Patriksson
Teacher on duty:	Peter Lindroth (0762-721860)
Result announced:	07–03–30 Short answers are also given at the end of the exam on the notice board for optimization in the MV building.

Exam instructions

When you answer the questions

Use generally valid theory and methods.
State your methodology carefully.

Only write on one page of each sheet. Do not use a red pen.
Do not answer more than one question per page.

At the end of the exam

Sort your solutions by the order of the questions.
Mark on the cover the questions you have answered.
Count the number of sheets you hand in and fill in the number on the cover.

Question 1

(LP duality)

Consider the linear programming problem to

$$\begin{array}{lrrrrrrl} \text{minimize} & z = & 2x_1 & +x_2 & +\alpha x_3 & & +x_5, & \\ \text{subject to} & & x_1 & +2x_2 & +3x_3 & -x_4 & +3x_5 & \geq 3, \\ & & -2x_1 & +x_2 & & +3x_4 & -2x_5 & \geq 4, \\ & & x_1, & x_2, & x_3, & x_4, & x_5 & \geq 0. \end{array}$$

(2p) a) Let $\alpha = 1$. Solve the resulting problem, *without using the Simplex method*, and state both $\boldsymbol{x}^*$ and z^*. Motivate your answer!

Hint: if you, for some reason, would like to solve some related problem with fewer variables, you are allowed to do it graphically.

(1p) b) Motivate, using a graph of the dual problem, the interval of α around $\alpha = 1$ for which $\boldsymbol{x}^*$ from a) remains optimal.

(3p) Question 2

(sufficiency of the KKT conditions under convexity)

Consider the problem to find

$$\begin{aligned} f^* := \underset{x}{\text{infimum}} \ & f(\boldsymbol{x}), \\ \text{subject to} \ & g_i(\boldsymbol{x}) \leq 0, \qquad i = 1, \dots, m, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, m$, are given differentiable and convex functions. State the KKT conditions for this problem, and assume that a vector $\boldsymbol{x}^*$ satisfies them. Establish that $\boldsymbol{x}^*$ then is a global optimum.

(3p) Question 3

(Farkas Lemma)

Consider the linear optimization problem to

$$\begin{array}{lrl} \text{minimize} & z = & x_2 - x_3 + 2x_4 + x_5 + 3x_6, \\ \text{subject to} & & x_1 + x_2 - 2x_3 + x_4 + 2x_6 \geq 0 \\ & & -x_1 + x_3 + x_4 + x_5 + x_6 \geq 0. \end{array}$$

Using Farkas Lemma, show that $z \geq 0$ holds for all feasible solutions.

(3p) Question 4

(the Frank–Wolfe algorithm)

As applied to the problem of minimizing a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ over a non-empty and bounded polyhedral set $X \subset \mathbb{R}^n$, the Frank–Wolfe method is defined, in short, thus: provide a first feasible solution $\boldsymbol{x}_0$ to the problem, and let $k := 0$; for given $\boldsymbol{x}_k$, solve the LP problem to minimize $\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{y}$ over $\boldsymbol{y} \in X$, and let $\boldsymbol{y}_k$ be an optimal solution to this problem. If the value of $\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{y}_k - \boldsymbol{x}_k)$ is (near) zero, then terminate with $\boldsymbol{x}_k$ being a (near-)stationary point, otherwise let $\boldsymbol{p}_k := \boldsymbol{y}_k - \boldsymbol{x}_k$ and perform a line search in the value of f along the direction $\boldsymbol{p}_k$ from $\boldsymbol{x}_k$, with a maximum step length of 1. Let the resulting vector be $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$, where α_k is the step length obtained in the line search. Let finally $k := k + 1$, and repeat.

Consider the nonlinear program to

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) := 10(x_1 + 1)^2 + (x_2 - 1)^2 \\ \text{subject to} & \boldsymbol{x} \in X := \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 1, \quad 0 \leq x_2 \leq 2 \,\}. \end{array}$$

Starting at $\boldsymbol{x}^0 = (1, 2)^{\mathrm{T}}$, solve this problem by using the Frank–Wolfe method using exact line searches. For each iteration, provide the best lower and upper bounds on the optimal value f^* of f. Motivate the termination of the algorithm.

Question 5

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

Given is the problem to minimize the function $f : \mathbb{R}^n \to \mathbb{R}$ over $\mathbb{R}^n$. We suppose that f is in C^2.

(1p) a) First, derive the basic Newton method with line searches.

The Levenberg–Marquardt modification of the Newton method is based on the possible failure of the Hessian matrix to be positive definite. Let $\boldsymbol{x} \in \mathbb{R}^n$. The search direction provided by the Levenberg–Marquardt modification of the Newton method is the solution to the equation

$$[\nabla^2 f(\boldsymbol{x}) + \mu \boldsymbol{I}^n]\boldsymbol{p} = -\nabla f(\boldsymbol{x}),$$

where $\mu \geq 0$ is chosen such that the eigenvalues of the matrix $\nabla^2 f(\boldsymbol{x}) + \mu \boldsymbol{I}^n$ is positive definite.

Consider now the following trust region problem:

$$\begin{aligned} \text{minimize} \quad & g(\boldsymbol{p}) := \frac{1}{2}\boldsymbol{p}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{p} + \boldsymbol{p}^{\mathrm{T}} \nabla f(\boldsymbol{x}), \\ \text{subject to} \quad & \|\boldsymbol{p}\|^2 \leq \delta, \end{aligned}$$

where $\delta > 0$. Show that the optimal solution is equivalent to a Levenberg–Marquardt step p with shift parameter 2μ, where μ is the Lagrange multiplier for the constraint.

(1p) b) Consider the following nonlinear problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(\boldsymbol{x}) := x_1^2 \left(\frac{x_1^2}{4} - \frac{2x_1}{3} - \frac{3}{2} \right) + (x_2 - 1)^2.$$

Start at $x^0 = (1, 0)^{\mathrm{T}}$ and perform one iteration with the Levenberg–Marquardt method.

(1p) c) There are algorithms for approximately solving trust region problems. Some of them are based on using the "Cauchy step". The Cauchy step is the solution to the trust region problem above, with the additional restriction that $\boldsymbol{p} \in \operatorname{span}\{\nabla f(\boldsymbol{x})\}$ (that is, that $\boldsymbol{p} = \alpha \nabla f(\boldsymbol{x})$ for some $\alpha \in \mathbb{R}$). Assume that the Hessian $\nabla^2 f(\boldsymbol{x})$ is positive definite. Compute the Cauchy step.

(3p) Question 6

(modelling)

In this problem, your task is to model the steel production strategy for a fictive company as a linear program. The problem is a simplified version of an old project assignment.

To produce steel, coal and iron ore are needed. Both these raw materials are taken from mines. There is one coal mine and three geographically separated ore mines available. There are two mills where steel is produced using the raw materials. At the mills the steel is formed into two types of products, plates and pipes. These products are then sold to the market.

The cost of mining coal is \$$g$/ton and the transport cost from the mine to each of the mills is \$$r_j$/ton, $j = 1, 2$. There is no limitation on the amount of coal that can be mined. The cost of mining iron ore is \$$h$/ton, the same for all mines. The tranport costs from mine i to mill j is \$$t_{ij}$/ton, $i = 1, 2, 3; j = 1, 2$. The maximum amount of iron ore that can be mined from mine i is cp_i tonnes.

To produce one ton of steel, a tonnes of coal, b tonnes of iron ore and c kWh of energy are needed. The cost of the energy is \$$p$/kWh and other process costs are \$$q$/ton of produced steel. One ton of steel can then be used to produce e_1 plates or e_2 pipes. The customers pick up the plates and pipes at the mills, and the prices are \$$s_1$/plate and \$$s_2$/pipe respectively. An estimation has been done that in total, one will not be able to sell more than d_1 plates and d_2 pipes.

You can handle all amounts as being continuous. Formulate the problem of maximizing the profit as a linear program. Define your variables carefully and explain your constraints. If you use indices for your constraints, indicate for which values they are used.

Question 7

(linear programming duality and optimality)

Let $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, and consider the canonical LP problem

$$\begin{aligned} \text{minimize} \quad & z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}, \\ & \boldsymbol{x} \geq \boldsymbol{0}^n. \end{aligned}$$

We denote the problem by (P).

(1p) a) Formulate explicitly the Lagrangian dual problem corresponding to the Lagrangian relaxation of *all* constraints of (P). (That is, the dimension of the Lagrangian dual problem is $m + n$.) Establish that this Lagrangian dual problem is equivalent to the canonical LP dual (D) of (P).

(2p) b) In the context of Lagrangian duality in nonlinear programming, the standard formulation of the primal problem is that to find

$$\begin{aligned} f^* := \underset{x}{\text{infimum}} \; & f(\boldsymbol{x}), & & (1) \\ \text{subject to} \; & g_i(\boldsymbol{x}) \leq 0, \qquad i = 1, \dots, \ell, \\ & \boldsymbol{x} \in X, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, 2, \dots, \ell)$ are given functions, and $X \subseteq \mathbb{R}^n$.

Identify the LP problem (P) as a special case of the general problem (1). State the global optimality conditions for the problem (1) and establish that when applied to the problem (P) they are equivalent to the primal–dual optimality conditions for the primal–dual pair (P), (D) of LP problems.

Good luck!

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:

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

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) $z^* = \frac{13}{9} = w^*$.

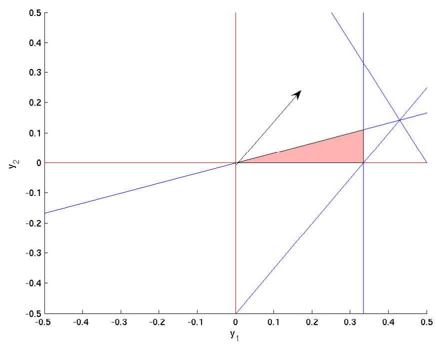

Figure 1: The dual feasible set.

(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

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

(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

$$\begin{aligned}
\sum_j y_{ij} &\leq cp_i, && i = 1, 2, 3, && (1)\\
\sum_j w_{jk} &\leq d_k, && k = 1, 2, && (2)\\
v_j &\leq \frac{x_j}{a}, && j = 1, 2, && (3)\\
v_j &\leq \sum_i \frac{y_{ij}}{b}, && j = 1, 2, && (4)\\
v_j &\leq \frac{u_j}{c}, && j = 1, 2, && (5)\\
v_j &\geq \sum_k \frac{w_{jk}}{e_k}, && j = 1, 2, && (6)\\
x_j, y_{ij}, u_j, v_j, w_{jk} &\geq 0, && \forall i, j, k, && (7)
\end{aligned}$$

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.