Chalmers/GU
Mathematics sciences

EXAM

TMA947/MMG621
NONLINEAR OPTIMISATION

Date:	23–10–26
Time:	8^{30}–13^{30}
Aids:	Chalmers approved calculator
Number of questions:	7; a passed question requires 2 points of 3.
	Questions are *not* numbered by difficulty.
	To pass requires 10 points and three passed questions.
Examiner:	Axel Ringh (073 708 23 73 and/or 031 772 12 34)

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.

Question 1

(LP duality)

Consider the linear programming problem

$$
\begin{aligned}
\text{minimize} \quad & 13x_1 + 14x_2 + 15x_3 + 16x_4 \\
\text{subject to} \quad & x_1 + 2x_2 + 3x_3 + 4x_4 \leq 1 \\
& 5x_1 + 6x_2 + 7x_3 + 8x_4 \geq 2 \\
& 9x_1 + 10x_2 + 11x_3 + 12x_4 = 3 \\
& x_1, x_2, x_3, x_4 \geq 0.
\end{aligned}
$$

(2p) a) State the dual problem.

(1p) b) An optimal solution to the primal problem is

$$
x^* = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.25 \end{bmatrix}.
$$

Give an optimal solution to the dual problem.

Question 2

(Separation Theorem)

The separation theorem is an important theorem in convex analysis and optimization. Somewhat informally, it states that for a closed and convex set $C \subset \mathbb{R}^n$ which is nonempty, if we take a point that is not in C, we can find a hyperplane that separates the space into two halfspaces so that the point is on one side of the hyperplane and C is on the other side of the hyperplane.

(1p) a) Give the formal statement of the separation theorem.

(2p) b) Prove the separation theorem. Do so using basic results from the course. If you rely on other results when performing your proof of the theorem, then those results must be stated explicitly; they may however be utilized without proof.

(3p) Question 3

(Frank-Wolfe method)

Your not so reliable friend AR wants to solve an optimization problem of the form

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & x_1 + \frac{1}{2}x_2 \geq 2 \\
& x_1 - \frac{1}{3}x_2 \geq \frac{1}{3} \\
& x_1 + x_2 \leq 7 \\
& x_1 \geq 0 \\
& x_2 \geq 0,
\end{aligned}
$$

for some continously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$. To do so, AR has implemented the problem in a computer and applied an iterative algorithm. The output from the first few steps of the algorithm is displayed in the table below.

k	x^k	$\nabla f(x^k)$	p^k	α_k	x^{k+1}
0	$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 2 \end{bmatrix}$	$\frac{1}{2}$	$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$
1	$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -2 \\ -2 \end{bmatrix}$	$\frac{1}{2}$	$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$
2	$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -3 \end{bmatrix}$	$\frac{1}{2}$	$\begin{bmatrix} 2 \\ \frac{3}{2} \end{bmatrix}$
$\vdots$					

In this table, x^k is the current iterate (point), $\nabla f(x^k)$ is the gradient in that point, p^k is the search direction used from the current iterate, α_k is the step length, and $x^{k+1} = x^k + \alpha_k p^k$ is the next iterate. AR knows that the step length computations are done correctly using the Armijo step length rule, however AR has forgotten which algorithm that was used to compute the feasible descent directions. Therefore, AR has a hard time reproducing the results. Help AR by verifying that the steps displayed in the table are consistent with (that is, can be the output of) the Frank-Wolfe method.

Question 4

(True or False)

The below three claims should be assessed. For each claim: Clearly state whether it is true or false. Provide an answer together with a short but complete motivation.

(1p) a) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x) = -x_1^4 + x_1^2 x_2^2 + 0.5x_2^3 - x_1 + 3,$$

and consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \mathbb{R}^2. \end{aligned}$$

Assume that we want to solve this problem using Newton's method with the Levenberg-Marquardt modification. The initial point is $x^0 = [1, 1]^T$.

Claim: The parameter $\gamma = 10.5$ is a valid modification parameter for Newton's method with the Levenberg-Marquardt modification in the point x^0.

(1p) b) Let $g_1 : \mathbb{R}^3 \to \mathbb{R}$, $g_2 : \mathbb{R}^3 \to \mathbb{R}$, and $g_3 : \mathbb{R}^3 \to \mathbb{R}$ be defined by

$$\begin{aligned} g_1(x) &= x_1^2 + x_2^2 - 4, \\ g_2(x) &= x_2^2 + x_3^2 - 9, \\ g_3(x) &= -x_1^3 + x_2 + x_3, \end{aligned}$$

respectively, and let $S = \{x \in \mathbb{R}^3 \mid g_i(x) \leq 0 \text{ for } i = 1, 2, 3\}$.

Claim: The set S is not convex.

(1p) c) Let $g_1 : \mathbb{R}^2 \to \mathbb{R}$ and $g_2 : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$\begin{aligned} g_1(x) &= (x_1 - 1)^2 + x_2^2 - 1, \\ g_2(x) &= (x_1 + 1)^2 + x_2^2 - 1 \end{aligned}$$

respectively, and let $S = \{x \in \mathbb{R}^2 \mid g_i(x) \leq 0 \text{ for } i = 1, 2\}$.

Claim: The vector $\bar{p} = [1,\ 0]^T$ is an element in the tangent cone in the point $\bar{x} = [0,\ 0]^T$. That is, $\bar{p} \in T_S(\bar{x})$.

(3p) Question 5

(Unconstrained optimization)

Let N be a positive integer, and let $a^{(i)} \in \mathbb{R}^m$ for $i = 1, 2, \ldots, N$ be given vectors. For these N vectors of dimension m, we define the mean value as

$$\bar{a} = \frac{1}{N} \sum_{i=1}^{N} a^{(i)}.$$

Show that $\bar{a}$ is the unique global minimizer of the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \sum_{i=1}^{N} \|x - a^{(i)}\|^2 \\ \text{subject to} \quad & x \in \mathbb{R}^m. \end{aligned}$$

Question 6

(the Karush-Kuhn-Tucker conditions)

Consider the problem

$$\begin{cases} \text{minimize} & f(x), \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, 2 \\ & x \in \mathbb{R}^2, \end{cases}$$

where

$$\begin{aligned} f(x) &= -x_1^2 - (x_2 - 2)^2, \\ g_1(x) &= x_1^2 - x_2, \\ g_2(x) &= -x_1 - 1. \end{aligned}$$

(1p) a) State the Karush-Kuhn-Tucker (KKT) conditions for the problem.

(1p) b) Find all KKT points, i.e., all feasible points x for which there is a solution to the KKT-system. Solutions based on graphical considerations are allowed, but they need to be supplemented with exact mathematical expressions and calculations motivating the conclusions.

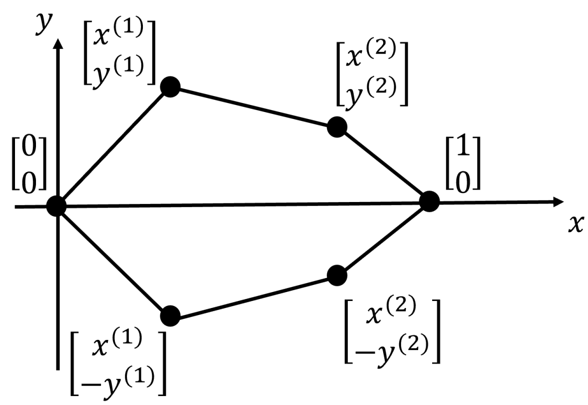
Hint: Consider the point $x = [\sqrt{3/2},\, 3/2]^T$.

(1p) c) Which of the KKT points has the smallest objective function value? Is this KKT point globally optimal?

(3p) Question 7

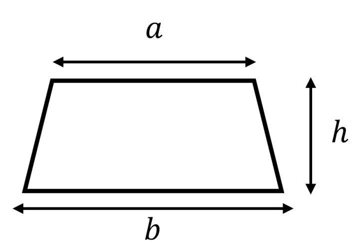
(Modelling)

Consider the hexagon shown in the image below. Formulate an optimization problem for finding the values of $x^{(1)}$, $y^{(1)}$, $x^{(2)}$ and $y^{(2)}$ that maximize the area of the hexagon in such a way that the diameter of the hexagon is less than or equal to 1. The latter means that all corner points of the hexagon must be at a distance of at most 1 from all other corner points.



Hint 1: To keep the ordering of the corner points as in the figure, note that one must have $x^{(1)} \leq x^{(2)}$, $y^{(1)} \geq 0$, and $y^{(2)} \geq 0$.

Hint 2: The area of a trapezoid with sides as in the image below is $A = \frac{1}{2}(a + b)h$.



Chalmers/GU
Mathematics

EXAM SOLUTION

TMA947/MMG621
NONLINEAR OPTIMISATION

Date: 23–10–26
Examiner: Axel Ringh

Note: the solutions presented here are brief in relation to the requirements on your answers, in particular regarding your motivations.

(3p) Question 1

(LP duality)

(2p) a)

$$\begin{aligned} \text{maximize} \quad & y_1 + 2y_2 + 3y_3 \\ \text{subject to} \quad & y_1 + 5y_2 + 9y_3 \leq 13 \\ & 2y_1 + 6y_2 + 10y_3 \leq 14 \\ & 3y_1 + 7y_2 + 11y_3 \leq 15 \\ & 4y_1 + 8y_2 + 12y_3 \leq 16 \\ & y_1 \leq 0 \\ & y_2 \geq 0 \\ & y_3 \in \mathbb{R}. \end{aligned}$$

(1p) b) The optimal primal objective function value is $f^* = 4$, so the dual objective function value is bounded from above by 4 on the dual feasible region. A couple of feasible points in the dual that achieve this objective function value are $y = [4, 0, 0]^T$, $y = [0, 2, 0]^T$, and $y = [0, 0, 4/3]^T$, which hence must all be optimal solutions.

(Alternatively: From the primal optimal solution and complementary in linear programming, we know that a dual optimal solution must be tight on the forth constraint, i.e., the inequality must be satisfied with equality. Moreover, we see that $4y_1 + 8y_2 + 12y_3 \leq 16$ is equivalent with $y_1 + 2y_2 + 3y_3 \leq 4$, which means that the objective function is bounded from above by 4. Therefore, any point in the dual that satisfies the last inequality with equality and that also satisfies the other three inequalities will be optimal. One such point is $y^* = [4, 0, 0]^T$.)

Question 2

(Separation Theorem)

(1p) a) See Theorem 4.29 in the book.

(2p) b) See the proof of Theorem 4.29 in the book.

(3p) Question 3

(Frank-Wolfe method)

To be the output of the Frank-Wolfe method, the search directions should be of the form $p^k = y^k - x^k$ where y^k is an optimal solution to

$$\begin{aligned} \text{minimize} \quad & \nabla f(x^k)^T z \\ \text{subject to} \quad & z_1 + \frac{1}{2} z_2 \geq 2 \\ & z_1 - \frac{1}{3} z_2 \geq \frac{1}{3} \\ & z_1 + z_2 \leq 7 \\ & z_1 \geq 0 \\ & z_2 \geq 0. \end{aligned}$$

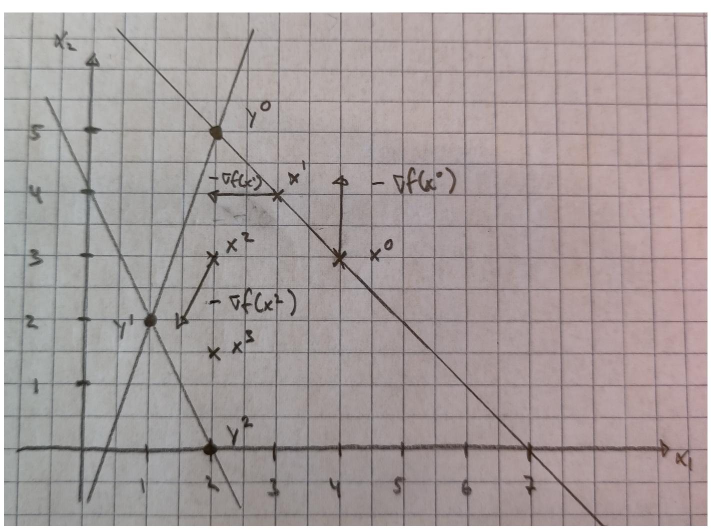
Writing the problem in standard form gives

$$\begin{aligned} \text{minimize} \quad & \nabla f(x^k)^T z \\ \text{subject to} \quad & Az = b \\ & z_1, z_2, s_1, s_2, s_3 \geq 0, \end{aligned}$$

where

$$A = \begin{bmatrix} 1 & \frac{1}{2} & -1 & 0 & 0 \\ 1 & -\frac{1}{3} & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 2 \\ \frac{1}{3} \\ 7 \end{bmatrix}.$$

We can now verify that $y^0 = x^0 + p^0 = [2,\ 5]^T$, $y^1 = x^1 + p^1 = [1,\ 2]^T$, and $y^2 = x^2 + p^2 = [2,\ 0]^T$ are indeed optimal solutions to the corresponding problems. The latter can be done in many different ways, e.g., by investigating which constraints an optimal point y^k is tight on, forming the corresponding BFS, and verifying that the BFS is optimal by computing the reduced costs. A graphical sketch is given in the picture below.



Question 4

(True or False)

The below three claims should be assessed. For each claim: clearly state whether it is true or false. Provide an answer together with a short but complete motivation.

(1p) a) False. The Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} -12x_1^2 + 2x_2^2 & 4x_1x_2 \\ 4x_1x_2 & 3x_2 + 2x_1^2 \end{bmatrix}$$

and thus

$$\nabla^2 f(x^0) + \gamma I = \begin{bmatrix} -10 & 4 \\ 4 & 5 \end{bmatrix} + 10.5 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 4 \\ 4 & 15.5 \end{bmatrix}.$$

The eigenvalues of this matrix are $\lambda_1 = -0.5$ and $\lambda_2 = 16.5$, which means that it is not positive definite (there are also other ways to check that it is not positive definite, e.g., the determinant is negative).

(1p) b) True. The two points $x^1 = [1, 0, 1]^T$ and $x^2 = [0, 0, 0]^T$ both belong to the set (verify the constraints!), but the point $\bar{x} = \frac{1}{2}x^1 + \frac{1}{2}x^2 = [\frac{1}{2}, 0, \frac{1}{2}]^T$ does not (violates constraint g_3).

(1p) c) False. $S = \{[0, 0]^T\}$, so $T_S(\bar{x}) = \{[0, 0]^T\}$ and thus $\bar{p} \notin T_S(\bar{x})$. (Other way to see it: $G(\bar{x}) = \{p \in \mathbb{R}^2 \mid p_1 = 0\}$ and $T_S(\bar{x}) \subset G(\bar{x})$, but $\bar{p} \notin G(\bar{x})$).

(3p) Question 5

(Unconstrained optimization)

Denoting the cost function $f(x)$, the gradient is given by

$$\nabla f(x) = \sum_{i=1}^{N}(x - a_i) = Nx - \sum_{i=1}^{N} a_i.$$

Moreover, the Hessian is given by

$$\nabla^2 f(x) = NI \succ 0 \quad \text{for all } x \in \mathbb{R}^m,$$

where $I \in \mathbb{R}^{m\times m}$ is the identity matrix. This means that f is a strictly convex function (motivate! Theorem 3.49). If the unconstrained optimization problem has an optimal solution, the latter is unique (motivate! Proposition 4.10). Moreover, $x^\star$ is a globally optimal solution to the optimization problem if and only if $\nabla f(x^\star) = 0$ (motivate! Theorem 4.18). Now,

$$\nabla f(x^\star) = 0 \quad \Longleftrightarrow \quad x^\star = \frac{1}{N}\sum_{i=1}^{N} a_i = \bar{a}.$$

Question 6

(the Karush-Kuhn-Tucker conditions)

(1p) a) For feasible points x, i.e., x such that $g_i(x) \leq 0$, $i = 1, 2$, the KKT conditions are

$$\nabla f(x) + \sum_{i=1}^{2} \mu_i \nabla g_i(x) = \begin{bmatrix} -2x_1 \\ -2(x_2 - 2) \end{bmatrix} + \mu_1 \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix} + \mu_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\mu_1, \mu_2 \geq 0,$$
$$\mu_1(x_1^2 - x_2) = 0, \quad \mu_2(-x_1 - 1) = 0$$

(1p) b) Drawing the problem guides us to considering the points $[0,\ 2]^T$, $[-1,\ 2]^T$, $[0,\ 0]^T$, and $[\sqrt{3/2},\ 3/2]^T$ (using the hint).

- For $[0,\ 2]^T$, no constraint is active but the gradient of f is is zero.
- For $[-1,\ 2]^T$, we find $\mu_1 = 0$ (g_1 not active) and $\mu_2 = 2$.
- For $[0,\ 0]^T$, we find $\mu_1 = 4$ and $\mu_2 = 0$ (g_2 not active).
- For $[\sqrt{3/2},\ 3/2]^T$, we find $\mu_1 = 1$ and $\mu_2 = 0$ (g_2 not active).

(1p) c) The smallest value among the KKT points is obtained at the point $\hat{x} = [0,\ 0]^T$. But f is unbounded from below on the feasible region, so $\hat{x}$ is not globally optimal.

(3p) Question 7

(Modelling)

For the hexagon in the below image, the area is given by the expression

$$\underbrace{\tfrac{1}{2}2y^{(1)}x^{(1)}}_{\text{left triangle}}+\underbrace{\tfrac{1}{2}(2y^{(1)}+2y^{(2)})(x^{(2)}-x^{(1)})}_{\text{trapezoid}}+\underbrace{\tfrac{1}{2}2y^{(2)}(1-x^{(2)})}_{\text{right triangle}}$$

To make sure that no two points are further away than distance 1, we need the constraints

$$\begin{aligned}
\sqrt{(x^{(1)}-0)^2+(y^{(1)}-0)^2} = \sqrt{(x^{(1)})^2+(y^{(1)})^2} &\le 1\\
\sqrt{(x^{(1)}-x^{(1)})^2+(y^{(1)}-(-y^{(1)}))^2} = 2y^{(1)} &\le 1,\\
\sqrt{(x^{(1)}-x^{(2)})^2+(y^{(1)}-(-y^{(2)}))^2} = \sqrt{(x^{(1)}-x^{(2)})^2+(y^{(1)}+y^{(2)})^2} &\le 1,\\
\sqrt{(x^{(1)}-1)^2+(y^{(1)}-0)^2} = \sqrt{(x^{(1)}-1)^2+(y^{(1)})^2} &\le 1\\
\sqrt{(x^{(2)}-1)^2+(y^{(2)}-0)^2} = \sqrt{(x^{(2)}-1)^2+(y^{(2)})^2} &\le 1\\
\sqrt{(x^{(2)}-x^{(2)})^2+(y^{(2)}-(-y^{(2)}))^2} = 2y^{(2)} &\le 1,\\
\sqrt{(x^{(2)}-0)^2+(y^{(2)}-0)^2} = \sqrt{(x^{(2)})^2+(y^{(2)})^2} &\le 1.
\end{aligned}$$

The solution is to maximize the expression for the area, where $x^{(1)}$, $y^{(1)}$, $x^{(2)}$ and $y^{(2)}$ are the variables, subject to the above constraints and $x^{(1)} \le x^{(2)}$, $y^{(1)} \ge 0$, and $y^{(2)} \ge 0$.