Chalmers/GU
Mathematics

EXAM

TMA947/MMG621
NONLINEAR OPTIMISATION

Date:	18–04–05
Time:	8^{30}–13^{30}
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:	Quanjiang Yu, tel. 0764-147839
Result announced:	18–05–03 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

(the simplex method)

Consider the following linear program:

$$\begin{aligned}
\text{maximize} \quad & z = x_1 + 2x_2, \\
\text{subject to} \quad & x_1 + x_2 \geq -1, \\
& x_1 - x_2 \geq 1, \\
& x_1, \quad x_2 \geq 0.
\end{aligned}$$

(2p) a) Solve the problem using phase I and phase II of the simplex method.

Aid: You may utilize the identity

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(1p) b) If an optimal solution exists, then use your calculations to decide whether it is unique or not. If the problem is unbounded, then use your calculations to specify a direction of unboundedness of the objective value.

(3p) Question 2

(gradient projection)

The gradient projection algorithm is a generalization of the steepest descent method to problems defined over convex sets. Given a point $\boldsymbol{x}_k$ the next point is obtained according to $\boldsymbol{x}_{k+1} = \text{Proj}_X[\boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k)]$, where X is the convex set over which we minimize, $\alpha_k > 0$ is the step length, and $\text{Proj}_X(\boldsymbol{y}) := \text{argmin}_{x \in X} ||\boldsymbol{x} - \boldsymbol{y}||$ (i.e., the closest point in X to $\boldsymbol{y}$). Note that if $X = \mathbb{R}$ then the method reduces to the method of steepest descent.

Consider the optimization problem to

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) := x_1^3 + 2x_2^2 + 2x_1x_2 - 2x_1, \\ \text{subject to} \quad & 0 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 2. \end{aligned}$$

Start at the point $\boldsymbol{x}_0 = (0, 2)^{\mathrm{T}}$ and perform one iteration of the gradient projection algorithm using step length $\alpha_k = 1/8$. Note that the special form of the feasible region X makes the projection very easy! Is the point obtained a global/local optimum? Motivate why/why not!

(3p) Question 3

(optimality conditions for special feasible sets)

Consider the problem of minimizing the function $f(\boldsymbol{x}) := \sum_{j=1,\ldots,n} f_j(x_j)$ over a set of the form $S := \{ \boldsymbol{x} \in \mathbb{R}^n \mid \sum_{j=1}^n x_j = b;\ x_j \geq 0, \forall j \}$. We assume that f is in C^1 on S, and of course that $b > 0$, such that S is non-empty.

This problem is often referred to as the *resource allocation problem*, since it entails allocating fractions of the resource b to "activity levels" x_j in an optimal manner, considering the minimization of the "cost function" f, and the available resources, represented by the value of b.

Utilize the optimality conditions for differentiable optimization over closed, convex sets to establish that any stationary point $\boldsymbol{x}^*$ must satisfy the conditions that for some value $\mu^* \in \mathbb{R}$ it holds that $f_j'(x_j^*) = \mu^*$, for all j with $x_j^* > 0$, while $f_j'(x_j^*) \geq \mu^*$, for all j with $x_j^* = 0$.

Question 4

(Karush-Kuhn-Tucker)

Consider the following problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) := 2x_1 - x_1^2, \\ \text{subject to} \quad & x_1^2 + x_2^2 \geq 25, \\ & x_1 \leq 4, \\ & x_2 \leq 4, \\ & x_1, x_2 \geq 0. \end{aligned}$$

(2p) a) State the KKT-conditions for the problem and check whether they are necessary or not, sufficient or not.

(1p) b) Find all KKT-points. Are the KKT points optimal? Motivate!

(3p) Question 5

(modelling)

The *set covering problem* is a classical question in combinatorics, computer science and complexity theory. Given a set of elements $\mathcal{U} = \{1, 2, ..., n\}$ (called the universe) and a collection $\mathcal{S}$ of m sets whose union equals the universe, the *set cover problem* is the problem to identify the smallest sub-collection of $\mathcal{S}$ whose union equals the universe.

For example, consider the universe $\mathcal{U} = \{1, 2, 3, 4, 5\}$ and the collection of sets $\mathcal{S} = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}\}$. Clearly the union of $\mathcal{S}$ is $\mathcal{U}$. However, we can cover all of the elements with the following, smaller number of sets: $\{\{1, 2, 3\}, \{4, 5\}\}$. This is also the smallest sub-collection whose union is $\mathcal{U}$.

A generalization of this problem is the *weighted set covering problem* where each set in $\mathcal{S}$ has a cost associated with it. The objective in the *weighted set covering problem* is to find a sub-collection of $\mathcal{S}$ whose union equals the universe, and so that the sum of the costs of the sets in the sub-collection is minimized.

Formulate an integer linear program (a linear objective function, linear constraints, and integrality restrictions on the variables) which models the weighted set covering problem.

Question 6

(true or false)

The below three claims should be assessed. Are they true or false, or is it impossible to say? Provide an answer, together with a short, but complete, motivation.

(1p) a) Suppose the function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a vector $\boldsymbol{x} \in \mathbb{R}^n$.

Claim: for the vector $\boldsymbol{p} \in \mathbb{R}^n$ to be a descent direction with respect to f at $\boldsymbol{x}$ it is necessary that $\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{p} < 0$.

(1p) b) Suppose you attack the problem of minimizing the twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ by means of Newton's method, using an exact line search. Suppose the iterate is $\boldsymbol{x}^t$, and that the result of iteration t is the next iterate $\boldsymbol{x}^{t+1}$.

Claim: $\nabla f(\boldsymbol{x}^{t+1})^{\mathrm{T}}(\boldsymbol{x}^{t+1} - \boldsymbol{x}^t) = 0$ holds.

(1p) c) *Claim:* A line segment in $\mathbb{R}^n$ is not a polyhedron.

(3p) Question 7

(Farkas' lemma)

Farkas' Lemma can be states as follows:

Let $\boldsymbol{A}$ be any $m \times n$ matrix and $\boldsymbol{b}$ an $m \times 1$ vector. Then exactly one of the two systems

$$\begin{aligned} \boldsymbol{A}x &= \boldsymbol{b}, \\ \boldsymbol{x} &\geq \mathbf{0}^n, \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} &\leq \mathbf{0}^m, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} &> 0, \end{aligned}$$

has a feasible solution, and the other system is inconsistent.

Prove Farkas' Lemma.

Chalmers/Gothenburg University
Mathematical Sciences

EXAM SOLUTION

TMA947/MMG621
NONLINEAR OPTIMISATION

Date: 18–01–09
Examiner: Michael Patriksson

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

Question 1

(the simplex method)

(2p) a) Rewrite the problem into standard form by adding/subtracting slack variables s_1 and s_2 to the left-hand side in the first and second constraint, respectively. Moreover, let $z := -z$ to get the problem on minimization form. Thus, we get the following linear program:

$$\begin{array}{llrrrrl}
\text{minimize} & z = & -x_1 & - & 2x_2, & & \\
\text{subject to} & & -x_1 & - & x_2 & +s_1 & = 1, \\
& & x_1 & - & x_2 & -s_2 & = 1, \\
& & x_1, & & x_2, & s_1, \quad s_2 & \geq 0.
\end{array}$$

Introducing the artificial variable a, phase I gives the problem

$$\begin{array}{llrrrrrl}
\text{minimize} & w = & a, & & & & & \\
\text{subject to} & & -x_1 & -x_2 & +s_1 & & & = 1, \\
& & x_1 & -x_2 & & -s_2 & +a & = 1, \\
& & x_1, & x_2, & s_1, & s_2, & a & \geq 0.
\end{array}$$

Using the starting basis $(s_1, a)^T$ gives

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{N} = \begin{pmatrix} -1 & -1 & 0 \\ 1 & -1 & -1 \end{pmatrix}, \boldsymbol{x}_B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{c}_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \boldsymbol{c}_N = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The reduced costs, $\bar{\boldsymbol{c}}_N^T = \boldsymbol{c}_N^T - \boldsymbol{c}_B^T \boldsymbol{B}^{-1} \boldsymbol{N}$, for this basis is $\bar{\boldsymbol{c}}_N^T = \begin{pmatrix} -1 & 1 & 1 \end{pmatrix}$, which means that x_1 enters the basis. The minimum ratio test implies that a leaves.

Thus, we move on to phase II using the basis $(s_1, x_1)^T$, and

$$\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \boldsymbol{N} = \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix}, \boldsymbol{x}_B = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \boldsymbol{c}_B = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \boldsymbol{c}_N = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

The new reduced costs are $\bar{\boldsymbol{c}}_N^T = \begin{pmatrix} -3 & -1 \end{pmatrix}$

which means that x_2 enters the basis. From the minimum ratio test we get $\boldsymbol{B}^{-1}\boldsymbol{N}_1 = \begin{pmatrix} -2 & -1 \end{pmatrix}^T < \boldsymbol{0}$, meaning that the problem is unbounded.

(1p) b) A direction of unboundedness is $\boldsymbol{l}(\mu) = \begin{pmatrix}1 & 0 & 2 & 0\end{pmatrix}^T + \mu\begin{pmatrix}1 & 1 & 2 & 0\end{pmatrix}^T$, $\mu \geq 0$.

(3p) Question 2

(gradient projection)

The gradient of f at the point $\boldsymbol{x}_0 = (0,2)^T$ is $\nabla f(\boldsymbol{x}_0) = (2,8)^T$. Taking a step in the negative gradient direction with $\alpha = 1/8$ gives the new point $\boldsymbol{x}_0 - (1/8)(2,8)^T = (-1/4, 1)$.

Projecting this point to the feasible set yields the new iterate $\boldsymbol{x}_1 = (0,1)$.

This point is clearly neither a local nor a global minimum. To check this, perform another iteration and see that the new iterate is not the same as $\boldsymbol{x}_1$.

(3p) Question 3

(optimality conditions for special feasible sets)

Thanks to the linearity of the constraints, the problem satisfies the Abadie constraint qualification and the Karush–Kuhn–Tucker conditions are necessary for the local optimality of $\boldsymbol{x}$. Introducing the multiplier μ for the equality constraint and λ_j for the sign constraints on x_j we obtain the Lagrangian function $L(\boldsymbol{x}, \mu, \boldsymbol{\lambda}) := b\mu + \sum_{j=1}^n (f_j(x_j) + [\mu - \lambda_j]x_j)$. Assume that $(\boldsymbol{x}^*, \mu^*, \boldsymbol{\lambda}^*)$ is a KKT point. Setting the partial derivatives of L to zero yields

$$\phi_j'(x_j^*) = \lambda_j^* - \mu^*, \qquad j = 1, \ldots, n, \tag{1}$$

and further, from complementarity, that

$$\lambda_j^* x_j^* = 0, \qquad j = 1, \ldots, n.$$

For a j with $x_j^* > 0$ it must then hold that $\phi_j'(x_j^*) = -\mu^*$. Suppose instead that $x_j^* = 0$. Then since $\lambda_j^* \geq 0$ must hold, we find, from the characterization (1), that $\phi_j'(x_j^*) = \lambda_j^* - \mu^* \geq -\mu^*$, and we are done.

Question 4

(Karush–Kuhn–Tucker)

(2p) a) Let $g_1(\boldsymbol{x}) := -x_1^2 - x_2^2 + 25$, $g_2(\boldsymbol{x}) := x_1 - 4$, $g_3(\boldsymbol{x}) := x_2 - 4$, $g_4(\boldsymbol{x}) := -x_1$ and $g_5(\boldsymbol{x}) := -x_2$with respective gradients $\nabla g_1 = (-2x_1, -2x_2)^T$, $\nabla g_2 = (1, 0)^T$,$\nabla g_3 = (0, 1)^T$, $\nabla g_4 = (-1, 0)^T$ and $\nabla g_5 = (0, -1)^T$.
Moreover, $\nabla f = (-2x_1 + 2, 0)^T$. The KKT-conditions are as follows:

$\nabla f(x^*) + \sum_{i=1}^{5} \mu_i \nabla g_i(x^*) = 0,$

$\mu_i g_i(x^*) = 0, i = 1, ..., 5,$

$\mu_i \geq 0, i = 1, ..., 5.$

Since the objective function f is not convex, the KKT conditions are not sufficient.

To prove KKT conditions are necessary, we use LICQ. For the interior points, there is no active constraints, and for the points on the boundary but not extreme points, there is only one active constraint, so the gradient of the active constraint must be linearly independent. So we only need to check three extreme points: $(4, 3)^T, (3, 4)^T, (4, 4)^T$. For point $(4, 3)^T$, the gradient of the active constraints are $(-8, -6)^T$ and $(1, 0)^T$, obviously they are linearly independent. For point $(3, 4)^T$, the gradient of the active constraints are $(-6, -8)^T$ and $(1, 0)^T$, obviously they are linearly independent. For point $(4, 4)^T$, the gradient of the active constraints are $(0, 1)^T$ and $(1, 0)^T$, obviously they are linearly independent. So LICQ holds at every feasible point. Thus, the KKT-conditions are necessary.

(1p) b) By letting different combinations of constraints be active, we can see when only g_2 active, we get $(4, a), 3 < a < 4$ are KKT points. When g_1 and g_2 are active, we get $(4, 3)$ is a KKT point. When g_2 and g_3 are active, we get $(4, 4)$ is a KKT point. So $(4, a), 3 \leq a \leq 4$ are KKT points. Since KKT conditions are necessary, so the optimal solution must be KKT points. Since all KKT points give the same objective function value -8, so all the KKT points are optimal.

(3p) Question 5

(modelling)

Let $S_1, S_2, \ldots, S_m$ be the sets, and let $U = \{1, \ldots n\}$ be the universe to cover. Now let the binary parameters $s_{ij} = 1$ if the element j is in the set S_i for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, and $s_{ij} = 0$ otherwise. Let w_i be the weight of set S_i.

Let x_i be a binary variable where $x_i = 1$ if set S_i is included in the sub-collection, where $i \in \{1, \ldots, m\}$, and $x_i = 0$ otherwise. The weighted set covering problem can now be formulated as:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{m} w_i s_i, \\
\text{subject to} \quad & \sum_{i=1}^{m} s_{ij} x_i \geq 1, \quad j \in \{1, \ldots, n\}, \\
& x_{ij} \in \{0, 1\}
\end{aligned}
$$

Question 6

(true or false)

(1p) a) False. Consider $f(x) = x^3$ at $x = 0$; a negative direction from 0 clearly reduces the value of f, while $f'(0) = 0$.

(1p) b) True. The claim is a characterization of the line search being exact in the direction of the vector $\boldsymbol{x}^{t+1} - \boldsymbol{x}^t$.

(1p) c) False. The solution set of the two linear inequalities $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geq b$ and $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leq b$, defines, by definition, a polyhedron, as it is the solution set of a collection of linear inequalities. On the other hand, the solution set also is a line segment in $\mathbb{R}^n$.

(3p) Question 7

(Farkas' lemma)

See the course book.