NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4180 Optimization I**

Academic contact during examination:
Phone:

Examination date: 27th May 2017
Examination time (from–to): 09:00–13:00
Permitted examination support material:

- The textbook: Nocedal & Wright, Numerical Optimization including errata.
- Rottmann, Mathematical formulae.
- Handouts on *Minimisers of Optimisation Problems*, *Basic Properties of Convex Functions*.
- Approved basic calculator.

Other information:

- All answers should be justified and include enough details to make it clear which methods or results have been used.
- You may answer to the questions of the exam either in English or in Norwegian.

Language: English
Number of pages: 11
Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date   Signature

Problem 1 Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = y^4 + 3y^2 - 4xy - 2y + x^2.$$

a) Compute all stationary points of f, and find all local or global minima of f.
(10 points)

We start by computing the gradient of f, which is

$$\nabla f(x,y) = \begin{pmatrix} -4y+2x \\ 4y^3+6y-4x-2 \end{pmatrix},$$

and set it to zero, that is, solve the system

$$\begin{aligned} -4y+2x &= 0, \\ 4y^3+6y-4x-2 &= 0. \end{aligned}$$

The first equation implies that $4x = 8y$; inserting this into the second equation yields

$$4y^3 - 2y - 2 = 0.$$

An obvious solution of this equation is $y = 1$; the corresponding x-coordinate is $x = 2$. Dividing the polynomial $4y^3-2y-2$ by $y-1$ now yields the equation

$$4y^2 + 4y + 2 = 0$$

for possible additional solutions. However, the only solutions of this equation are imaginary (namely $1/2 \pm i/2$). Thus the only stationary point of f is the point

$$(x,y) = (2,1).$$

Now we note that the function f is coercive, as we can write it as

$$f(x,y) = y^4 - y^2 - 2y + (x-2y)^2,$$

which obviously tends to infinity if $\|(x,y)\| \to \infty$. (If $y \to \pm\infty$, the term $y^4 - y^2 - 2y$ tends to ∞ and the term $(x-2y)^2$ is positive; thus the whole function tends to ∞. If y stays bounded but x tends to $\pm\infty$, the term $(x-2y)^2$ tends to ∞.) This implies that the function f has at least one global minimum, the only candidate for which is the only stationary point $(1,2)$. This implies that $(1,2)$ is a global (and thus also local) minimum.

b) Decide whether the function f is convex or not.
(5 points)

In order to check whether the function f is convex, we compute its Hessian as

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & -4 \\ -4 & 12y^2+6 \end{pmatrix}.$$

At the point $(x,y) = (0,0)$ we obtain

$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & -4 \\ -4 & 6 \end{pmatrix},$$

the determinant of which is negative. Thus $\nabla^2 f(0,0)$ is not positive semi-definite, and thus f is non-convex.

c) Starting at the point $(x,y) = (0,0)$, compute one step of the gradient descent method with backtracking (Armijo) linesearch (see Algorithm 3.1 in Nocedal and Wright). Start with an initial step length $\bar{\alpha} = 1$, and use the parameters $c = 1/10$ (sufficient decrease parameter) and $\rho = 1/4$ (contraction factor). *(10 points)*

The negative gradient of f at the point $(0,0)$ is

$$p := -\nabla f(0,0) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Moreover we have

$$f(0,0) = 0$$

and

$$p^T \nabla f(0,0) = -4.$$

Thus a step of length α in direction p is accepted by the line search, if

$$f(\alpha p) \le -\frac{4}{10}\alpha.$$

With $\alpha = 1$ we obtain

$$f(p) = f(0,2) = 24 \not\le -\frac{4}{10}.$$

Thus we decrease the step size to $\alpha = 1/4$. We check again and see that

$$f(p/4) = f(0,1/2) = -\frac{3}{16} \le -\frac{1}{10}.$$

Thus the step is accepted and the next iterate will be $(0, 1/2)$.

Problem 2 Consider the constrained optimization problem

$$2x^2 - y^2 - 2y \to \min \qquad \text{subject to } x + y = 1.$$

Formulate the unconstrained optimization problem resulting from an application of the quadratic penaly method applied to this problem. Determine for which parameters $\mu > 0$ the resulting unconstrained problems has a solution, and compute the solution for all parameters for which it exists.
(10 points)

Using the quadratic penalty method, we obtain the function

$$Q(x, y; \mu) = 2x^2 - y^2 - 2y + \frac{\mu}{2}(x + y - 1)^2$$

to be minimised instead of the original function. The gradient of Q is

$$\nabla Q(x, y; \mu) = \begin{pmatrix} 4x + \mu(x + y - 1) \\ -2y - 2 + \mu(x + y - 1) \end{pmatrix}.$$

Moreover, the Hessian of Q is

$$\nabla^2 Q(x, y; \mu) = \begin{pmatrix} 4 + \mu & \mu \\ \mu & -2 + \mu \end{pmatrix}.$$

Since $\mu > 0$, we always have that $4 + \mu > 0$. Moreover

$$\det \nabla^2 Q(x, y; \mu) = (4 + \mu)(\mu - 2) - \mu^2 = -8 + 2\mu,$$

and thus $\nabla^2 Q(x, y; \mu)$ is positive definite if and only if $\mu > 4$, and positive semi-definite if and only if $\mu \geq 4$. Since Q is quadratic, this implies that the function Q will have a minimiser if $\mu > 4$, but won't have a minimiser for $\mu < 4$. In case $\mu = 4$, we have a minimiser in case a stationary point exists.

Next we will try to compute stationary points, that is, set the gradient of Q to zero. From the equation

$$4x + \mu(x + y - 1) = -2y - 2 + \mu(x + y - 1)$$

we obtain that

$$y = -2x - 1.$$

Inserting this into the x-derivative of Q, we obtain the equation

$$0 = 4x + \mu(x - 2x - 1 - 1) = (4 - \mu)x - 2\mu.$$

Thus we don't have a stationary point for $\mu = 4$, and for $\mu \neq 4$ we have

$$x = \frac{2\mu}{4 - \mu}.$$

In this case,

$$y = -2x - 1 = -\frac{4\mu}{4-\mu} - 1 = \frac{3\mu+4}{\mu-4}.$$

Thus the minimiser of the quadratic penalty functional is $(-2\mu, 3\mu+4)/(4-\mu)$ for $\mu > 4$, and the functional does not have a minimizer for $\mu \le 4$.

Problem 3 We consider the problem of solving the constrained optimization problem

$$(x-2)^2 + (y-1)^2 \to \min \qquad \text{subject to } (x,y) \in \Omega,$$

where the set $\Omega \subset \mathbb{R}^2$ is given by the inequality constraints

$$\begin{aligned} y &\ge 0, \\ 1 - x &\ge 0, \\ x^2 - y &\ge 0. \end{aligned}$$

a) Sketch the set Ω and find all points $(x,y) \in \Omega$ where the LICQ fails to hold. *(5 points)*

(Sorry for being too lazy to sketch the set Ω...)

The constraints c_i describing the set Ω are

$$\begin{aligned} c_1(x,y) &= y, \\ c_2(x,y) &= 1 - x, \\ c_3(x,y) &= x^2 - y, \end{aligned}$$

with gradients

$$\begin{aligned} \nabla c_1(x,y) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ \nabla c_2(x,y) &= \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \\ \nabla c_3(x,y) &= \begin{pmatrix} 2x \\ -1 \end{pmatrix}. \end{aligned}$$

In particular, none of these gradients ever becomes zero, which implies that LICQ is satisfied whenever none (trivially) or at most one of the constraints is active. Also the gradients ∇c_1 and ∇c_2 are always linearly independent, and so are the gradients ∇c_2 and ∇c_3. Thus LICQ also holds if either c_1 and c_2, or c_2 and c_3 are simultaneously active. However, ∇c_1 and ∇c_3 are linearly dependent for $x = 0$. Thus LICQ fails to hold if c_1 and c_3 are active and $x = 0$. This is the case precisely for the point $(0,0)$. Finally, we see that it can never happen that all three constraints are simultaneously active.

In total, this implies that $(0,0)$ is the only point in Ω, where LICQ fails to hold.

b) Determine the tangent cone and the set of linearized feasible directions to the set Ω in the points $(x,y) = (1,1)$ and $(x,y) = (0,0)$.
(10 points)

In the point $(1,1)$, LICQ holds, and thus the tangent cone coincides with the set of linearised feasible directions, which is, as the constraints c_2 and c_3 are active in $(1,1)$, the set

$$\mathcal{F}(1,1) = \{d = (d_1,d_2) \in \mathbb{R}^2 : -d_1 \geq 0,\ 2d_1 - d_2 \geq 0\}.$$

In the point $(0,0)$, LICQ does not hold, and therefore we can only conclude that the tangent cone is contained in, but not necessarily equal to, the set of linearised feasible directions. Since the constraints c_1 and c_3 are active at $(0,0)$, the latter can be computed to

$$\mathcal{F}(0,0) = \{d = (d_1,d_2) \in \mathbb{R}^2 : d_2 \geq 0,\ -d_2 \geq 0\} = \mathbb{R} \cdot (1,0).$$

That is, the set of linearised feasible directions at $(0,0)$ is precisely the x-axis. However, since locally around $(0,0)$ the x-axis is actually contained in the set Ω, we can conclude that it is also contained in the tangent cone at $(0,0)$. Now this immediately implies that also the tangent cone at $(0,0)$ is equal to the x-axis, that is,

$$T_\Omega(0,0) = \mathbb{R} \cdot (1,0).$$

c) Use the second order optimality conditions in order to show that the point $(1,1)$ is a local solution of this optimization problem.
(15 points)

The KKT conditions for the optimisation problem at hand read as

$$\begin{aligned}
\begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix} &= \lambda_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 2x \\ -1 \end{pmatrix}, \\
y &\geq 0, \\
1 - x &\geq 0, \\
x^2 - y &\geq 0, \\
\lambda_i &\geq 0, \qquad i = 1,2,3, \\
\lambda_i c_i(x,y) &= 0, \qquad i = 1,2,3.
\end{aligned}$$

At the point $(1,1)$, the constraints are obviously satisfied, and the complementarity condition implies that the Lagrange parameter λ_1 has to be equal to zero (as the first constraint is inactive). Inserting $x = 1$ and $y = 1$ into the first condition, we thus obtain the equations

$$\begin{aligned} -2 &= -\lambda_2 + 2\lambda_3, \\ 0 &= -\lambda_3, \end{aligned}$$

for the (yet unknown) Lagrange parameters λ_2 and λ_3. We immediately see that these conditions imply that $\lambda_3 = 0$ and thus $\lambda_2 = 2$. Since both Lagrange parameters are non-negative, the KKT-conditions are satisfied with $\lambda_1 = \lambda_3 = 0$ and $\lambda_2 = 1$. Thus $(1, 1)$ is a KKT point.

Now we check the second order conditions for this point. The Lagrangian for our problem is

$$\mathcal{L}(x, y; \lambda_1, \lambda_2, \lambda_3) = (x - 2)^2 + (y - 1)^2 - \lambda_1 y - \lambda_2(1 - x) - \lambda_3(x^2 - y).$$

For $\lambda_1 = \lambda_3 = 0$ and $\lambda_2 = 2$, we obtain

$$\mathcal{L}(x, y; 0, 1, 0) = (x - 2)^2 + (y - 1)^2 - 2(1 - x).$$

In particular, the Hessian of this Lagrangian is

$$\nabla^2_{x,y}\mathcal{L}(x, y; 0, 1, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

which is obviously always positive definite, and thus in particular positive definite "on the critical cone." Thus the second order sufficient conditions for a (strict) local minimum hold at the point $(1, 1)$.

Problem 4 Find the dual of the linear optimization problem

$$x + y - z \to \min \qquad \text{subject to} \begin{cases} x - y - 3z = -1, \\ x,\ y,\ z \geq 0, \end{cases}$$

and compute the solutions of both the primal and the dual problem.
(10 points)

In general, the dual of a linear programme of the (standard) form

$$c^T x \to \min \qquad \text{subject to} \begin{cases} Ax = b, \\ x \geq 0 \end{cases}$$

is the programme

$$b^T \lambda \to \max \qquad \text{subject to } A^T \lambda \leq c.$$

In this case we have

$$c = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & -1 & -3 \end{pmatrix}, \qquad b = -1.$$

Thus we obtain the dual problem (for the *scalar* Lagrange parameter $\lambda \in \mathbb{R}$)

$$-\lambda \to \max \qquad \text{subject to } \lambda \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix} \leq \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

In other words, we want to minimise λ subject to the constraints

$$\lambda \leq 1, \qquad -\lambda \leq 1, \qquad -3\lambda \leq -1.$$

The obvious solution of this problem is

$$\lambda^* = 1/3.$$

In order to solve the primal problem, we observe that only the third constraint ($-3\lambda \leq -1$) was active at the dual solution $\lambda^* = 1/3$. Since the primal variables x^*, y^*, z^* are Lagrange parameters for the dual problem, this implies that $x^* = y^* = 0$. As a consequence, the equality constraint for the primal problem implies that $-3z^* = -1$ or $z^* = 1/3$. Thus the solution of the primal problem is

$$(x^*, y^*, z^*) = (0, 0, 1/3).$$

Problem 5 Assume that $C \subset \mathbb{R}^n$ is a closed and convex, non-empty set. Given any point $y \in \mathbb{R}^n$, we define the projection $P_C(y)$ of y to be the solution of the optimization problem

$$\min_x \frac{1}{2}(x - y)^2 \qquad \text{subject to } x \in C. \tag{1}$$

I sincerely apologise for the misprint in this exercise. The optimisation problem should of course be

$$\min_x \frac{1}{2}\|x - y\|^2 \qquad \text{subject to } x \in C.$$

In almost all answers I have received, this was interpreted correctly, though. In the few answers, where it was obviously misinterpreted, I have adjusted the grading scale.

a) Show that the optimization problem (1) has a solution, and that this solution is unique.
(5 points)

The function

$$g(x) = \frac{1}{2}\|x - y\|^2$$

is strictly convex and coercive (it is quadratic, and its Hessian is the identity matrix) and thus attains a minimiser over the closed set C. Since the set C is also convex, it follows that the minimiser is unique: Indeed, assume that x_1 and x_2 are two different solutions of (1). Then x_1, $x_2 \in C$ and $g(x_1) = g(x_2)$. Because of the convexity of C we have that $(x_1 + x_2)/2 \in C$, and because of the strict convexity of g we have that $g((x_1 + x_2)/2) < g(x_1)/2 + g(x_2)/2 = g(x_1) = g(x_2)$, which contradicts the minimality of x_1 and x_2. Thus the minimiser has to be unique.

We now assume that the set C is given by

$$x \in C \iff c(x) \geq 0,$$

where $c\colon \mathbb{R}^n \to \mathbb{R}$ is a concave (that is, $-c$ is convex) and smooth function.

b) Assume that there exists some $x \in C$ such that $c(x) > 0$. Show that the projection of a point y to the set C is characterized by the following conditions:

- If $y \in C$, then $P_C(y) = y$.
- If $y \notin C$, then $x = P_C(y)$, if and only if $x \in C$ and there exists $\lambda > 0$ such that

$$x - y = \lambda \nabla c(x).$$

(10 points)

Again, I have to apologise for a misprint in the problem setting, as the conditions in the second case should state that $x \in \partial C$ and $x - y = \lambda \nabla c(x)$ for some $\lambda > 0$ (that is, x has to lie on the boundary of C and not anywhere in C as claimed in the problem setting). I have adjusted the grading of this subproblem to account for my errors.

As a first step, we realise that the given conditions are nothing else than the KKT conditions for the optimisation problem (1) in disguise: The KKT conditions for (1) read as

$$\begin{aligned} x - y &= \lambda \nabla c(x), \\ c(x) &\geq 0, \\ \lambda c(x) &= 0. \end{aligned} \tag{2}$$

In case $y \in C$, they are obviously satisfied with $x = y$ and $\lambda = 0$. On the other hand, if $y \notin C$, then $x - y$ necessarily has to be different from 0 and thus the Lagrange parameter λ actually has to be strictly larger than 0, which in view of the complementarity condition $\lambda c(x) = 0$ implies that $c(x) = 0$. That is, for $y \notin C$, they can be written as

$$x - y = \lambda \nabla c(x) \text{ for some } \lambda > 0, \text{ and } c(x) = 0.$$

It remains to show that the KKT conditions are actually necessary and sufficient.

For the necessity of the KKT conditions, we have to show that LICQ holds everywhere in C, which follows from the existence of x with $c(x) > 0$ in the following way:
Since the function c is concave, its (global and local) maximisers are characterised by the condition $\nabla c(x) = 0$. By assumption, there exists x with $c(x) > 0$, which implies that no point x satisfying $c(x) = 0$ can be a maximiser of c. This in turn implies that $\nabla c(x) \neq 0$ whenever $c(x) = 0$, which means that LICQ holds everywhere. Thus the KKT conditions are necessary for a minimiser of (1).

For the sufficiency of the KKT conditions, we make use of the strict convexity of the problem, using for instance the following argumentation: The Hessian of the Lagrangian

$$\mathcal{L}(x; \lambda) = \frac{1}{2}\|x - y\|^2 - \lambda c(x)$$

of the problem we are solving is positive definite whenever $\lambda \geq 0$, because c is a concave function. Thus every solution of the KKT conditions is a strict local solution of the constrained optimisation problem. However, because of the convexity of the problem (that is, the function we are minimising and the set we are optimising over), every local solution is already a global solution. Thus any point satisfying the KKT conditions (2) is already a (the!) global solution of the problem, and thus precisely the projection $P_C(y)$.

We now consider the numerical solution of an optimization problem of the form

$$f(x) \to \min \qquad \text{subject to } x \in C, \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex and smooth function.

One method for solving this problem is the so called *gradient projection method*, which is defined by the iteration

$$x_{k+1} = P_C\Big(x_k - \tau \nabla f(x_k)\Big), \tag{4}$$

where $\tau > 0$ is some fixed parameter.

c) Assume that the iteration x_{k+1} converges to some $x^* \in \mathbb{R}^n$. Show that x^* is a solution of (3).
(10 points)

We note first that $x_k \in C$ for all k, and thus, as C is closed, also $x^* \in C$.
Assume now first that there is a subsequence $x_{k'}$ such that $x_{k'} - \tau\nabla f(x_{k'}) \in C$ for all k'. Then the definition of the projection implies that

$$x_{k'+1} = P_C(x_{k'} - \tau\nabla f(x_{k'})) = x_{k'} - \tau\nabla f(x_{k'}).$$

Since $\|x_{k'+1} - x_{k'}\| \to 0$ because of the convergence of the sequence x_k to x^*, this implies that $\nabla f(x_{k'}) \to 0$. Thus $\nabla f(x^*) = \lim_{k'\to\infty} \nabla f(x_{k'}) = 0$, implying that x^* is actually an unconstrained global minimum of f (that happens to lie in C).
Thus we may assume without loss of generality that $x_k - \tau\nabla f(x_k) \notin C$ for all sufficiently large k. In this case, we have that $c(x_k) = 0$ for all sufficiently large k and thus also $c(x^*) = 0$. Since c is concave and there exists x with $c(x) > 0$, this in particular implies that $\nabla c(x^*) \neq 0$. Moreover, by the characterisation of the projection there exist $\lambda_k > 0$ such that

$$x_{k+1} - x_k + \tau\nabla f(x_k) = \lambda_k \nabla c(x_k).$$

The left hand side of this equation tends to $\tau\nabla f(x^*)$ as $k \to \infty$, and we also have that $\nabla c(x_k) \to \nabla c(x^*) \neq 0$ as $k \to \infty$. As a consequence, the sequence λ_k necessarily has to converge to some $\lambda^* \geq 0$, and we have

$$\tau\nabla f(x^*) = \lambda^* \nabla c(x^*).$$

Since $c(x^*) = 0$ and $\lambda^* \geq 0$, this shows that the KKT conditions are satisfied at the point x^*.
Next we show that, because of the convexity of f and concavity of c, the KKT conditions are sufficient for a minimiser. To that end, we observe first that the Lagrangian

$$\mathcal{L}(x, \lambda^*) = f(x) - \lambda^* c(x)$$

is convex, as $\lambda^* \geq 0$. Since by construction

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

it follows that x^* is a global minimum of the Lagrangian (over the whole space $\mathbb{R}^n$). Now let $x \in C$ be arbitrary. Then the fact that $c(x) \geq 0$ together with the minimality of x^* implies that

$$f(x) \geq f(x) - \lambda^* c(x) = \mathcal{L}(x, \lambda^*) \geq \mathcal{L}(x^*, \lambda^*) = f(x^*) - \lambda^* c(x^*) = f(x^*).$$

In other words, $f(x) \geq f(x^*)$ for all $x \in C$, which shows that x^* is a solution of (3).

Remark: The proof becomes noticeably simpler, if one assumes/uses the continuity of the projection P_C (which actually holds). Then one may simply take the limit $k \to \infty$ in (4) *and immediately obtains the equation*

$$x^* = P_C(x^* - \tau \nabla f(x^*)),$$

which is the basis of all the subsequent considerations. In the grading of this problem, I was perfectly fine with answers that used this assumption.