NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4180 Optimization I**

Academic contact during examination:

Phone:

Examination date: 26th May 2016

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 unconstrained 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: 8

Number of pages enclosed: 0

Checked by:

Date Signature

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

$$f(x,y) = 2x^2 - 2xy^2 - 12x + y^4 + 2y^2 + 36.$$

a) Compute all stationary points of f and find all local or global minimizers of f.

We start by computing the gradient and the Hessian of f. We have

$$\nabla f(x,y) = \begin{pmatrix} 4x - 2y^2 - 12 \\ -4xy + 4y^3 + 4y \end{pmatrix}$$

and

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

The stationary points satisfy $\nabla f(x,y) = 0$. Now the second row of the gradient implies that

$$4y(-x + y^2 + 1) = 0$$

and thus either $y = 0$ or $y^2 = x - 1$. In the first case, the equation for the first row of the gradient implies that $x = 3$. Thus we have a stationary point

$$(x,y) = (3,0).$$

On the other hand, if $y^2 = x - 1$, then the equation for the first row of the gradient becomes

$$4x - 2(x-1) - 12 = 0 \qquad \text{or} \qquad x = 5,$$

which implies that in this case $y = \pm\sqrt{5-1} = \pm 2$. Thus we have the two stationary points

$$(x,y) = (5, \pm 2).$$

In order to determine whether they are local minima, we look at the Hessian of f. We have

$$\nabla^2 f(3,0) = \begin{pmatrix} 4 & 0 \\ 0 & -8 \end{pmatrix},$$

which is obviously indefinite. Thus $(3,0)$ is a saddle point of f. On the other hand

$$\nabla^2 f(5,\pm 2) = \begin{pmatrix} 4 & \mp 8 \\ \mp 8 & 48 \end{pmatrix}.$$

Since the diagonal entries are positive and the determinant of this matrix is also positive, the matrix is positive definite, which implies that the points $(5, \pm 2)$ are local minima.

Now note that we can write

$$f(x, y) = (x - y^2)^2 + (x - 6)^2 + 2y^2,$$

which is obviously coercive. Thus the function f attains at least one global minimum. Since f is symmetric with respect to y, it follows that $f(5, 2) = f(5, -2)$. Thus the function f has the same function value at the only two candidates for the global minimum, which implies that both of these points are actually global minima.

b) Starting at the point $(x, y) = (1, 2)$ compute one step of Newton's 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 $c = 1/8$ (sufficient decrease parameter) and $\rho = 1/2$ (contraction factor).

We need the gradient and the Hessian of f at $(1, 2)$. We have

$$\nabla f(1, 2) = \begin{pmatrix} -16 \\ 32 \end{pmatrix}$$

and

$$\nabla^2 f(1, 2) = \begin{pmatrix} 4 & -8 \\ -8 & 48 \end{pmatrix}.$$

Solving the system $\nabla^2 f(1, 2)p = -\nabla f(1, 2)$ yields the Newton direction

$$p = \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

The Armijo condition now requires that

$$f((1, 2) + \alpha p) \le f(1, 2) + \frac{1}{8}\alpha p^T \nabla f(1, 2) = 42 - 8\alpha.$$

With $\alpha = 1$ we have

$$f((1, 2) + p) = f(5, 2) = 10 \le 34 = 42 - 8\alpha.$$

Thus the step $\alpha = 1$ is accepted by the line search, and as a consequence the next iterate in the Newton method would be the point $(5, 2)$. (Since this is actually the global minimum, the iteration will stop there.)

Problem 2 Consider the constrained optimization problem

$$x^3 \to \min \qquad \text{subject to } x^2 + y^2 = 1. \tag{1}$$

Find (by whatever means) the solution of this problem. In addition, formulate the augmented Lagrangian for this problem and determine all parameters $\lambda \in \mathbb{R}$ and $\mu > 0$ for which the solution of (1) is a local minimizer of the augmented Lagrangian.

We want to minimize the function x^3 for (x, y) on the unit circle. Since $x \mapsto x^3$ is strictly increasing, it is obvious that the minimizer will be the point on the unit circle with minimal x-coordinate, which is the point $(-1, 0)$.

Now the augmented Lagrangian of the problem is defined as

$$\mathcal{L}_A(x, y; \lambda, \mu) := x^3 - \lambda(x^2 + y^2 - 1) + \frac{\mu}{2}(x^2 + y^2 - 1)^2.$$

In order to determine whether $(-1, 0)$ is a local minimizer of $\mathcal{L}_A$, we compute the gradient of $\mathcal{L}_A$. We have

$$\nabla\mathcal{L}_A(x, y; \lambda, \mu) = \begin{pmatrix} 3x^2 - 2x\lambda + 2x\mu(x^2 + y^2 - 1) \\ -2y\lambda + 2y\mu(x^2 + y^2 - 1) \end{pmatrix}$$

and thus

$$\nabla\mathcal{L}_A(-1, 0; \lambda, \mu) = \begin{pmatrix} 3 + 2\lambda \\ 0 \end{pmatrix}.$$

Thus $(-1, 0)$ is a stationary point of the augmented Lagrangian, if and only if $\lambda = -3/2$. Next we need the Hessian of $\mathcal{L}_A$, which is

$$\nabla^2\mathcal{L}_A = \begin{pmatrix} 6x - 2\lambda + 2\mu(x^2 + y^2 - 1) + 4\mu x^2 & 4\mu xy \\ 4\mu xy & -2\lambda + 2\mu(x^2 + y^2 - 1) + 4\mu y^2 \end{pmatrix}.$$

In particular, we have

$$\nabla^2\mathcal{L}_A(-1, 0; -3/2, \mu) = \begin{pmatrix} 4\mu - 3 & 0 \\ 0 & 3 \end{pmatrix},$$

which is positive definite for $\mu > 3/4$ and positive semi-definite for $\mu = 3/4$. As a consequence, $(-1, 0)$ is a local minimizer for the augmented Lagrangian if $\lambda = -3/2$ and $\mu > 3/4$ and is no local minimizer if either $\lambda \neq -3/2$ or $\lambda = -3/2$ and $\mu < 3/4$.

The only parameter values where it is not yet clear whether $(-1, 0)$ minimizes $\mathcal{L}_A$ are $\lambda = -3/2$ and $\mu = 3/4$. Here the augmented Lagrangian is

$$\mathcal{L}_A(x, y; -3/2, 3/4) = x^3 + \frac{3}{2}(x^2 + y^2 - 1) + \frac{3}{8}(x^2 + y^2 - 1)^2.$$

Moreover, for $y = 0$ we have

$$\mathcal{L}_A(x, 0; -3/2, 3/4) = \frac{3}{8}x^4 + x^3 + \frac{3}{4}x^2 - \frac{9}{8} =: p(x).$$

We have $p'(-1) = 0$ and $p''(-1) = 0$. However, $p'''(-1) \neq 0$, which implies that -1 is no local minimizer of p. As a consequence, $(-1, 0)$ is no local minimizer of $\mathcal{L}_A$ for $\lambda = -3/2$ and $\mu = 3/4$.

Problem 3 Consider the constrained optimization problem

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

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

$$xy \geq 0 \qquad \text{and} \qquad x - y(y + 2) \geq 0.$$

a) Sketch the set Ω and find all points $(x, y) \in \Omega$ for which the LICQ holds.

Denoting $c_1(x, y) = xy$ and $c_2(x, y) = x - y(y + 2)$, we have

$$\nabla c_1(x, y) = \begin{pmatrix} y \\ x \end{pmatrix} \qquad \text{and} \qquad \nabla c_2(x, y) = \begin{pmatrix} 1 \\ -2y - 2 \end{pmatrix}.$$

We consider first the points $(0, 0)$ and $(0, -2)$, where both constraints are active. At $(0, 0)$ we have $\nabla c_1(0, 0) = 0$, and thus the constraints are not linearly independent. At $(0, -2)$ we have $\nabla c_1(0, -2) = (-2, 0)$ and $\nabla c_2(0, -2) = (1, 2)$, which are obviously linearly independent. Thus the LICQ holds. Next we note that $\nabla c_1(x, y) \neq 0$ for $(x, y) \neq (0, 0)$, which implies that $\nabla c_1(x, y) \neq 0$ whenever only the constraint c_1 is active. Thus the LICQ holds on each of these points. Similarly, $\nabla c_2(x, y) \neq 0$ for all x and y, which implies that the LICQ holds whenever only the constraint c_2 is active. Finally, the LICQ trivially holds if no constraints are active.

Thus the LICQ holds everywhere in $\Omega \setminus \{(0, 0)\}$.

b) Determine the set $\mathcal{F}(0, 0)$ of linearized feasible direction at $(0, 0)$ and show that the tangent cone to Ω at $(0, 0)$ is different from $\mathcal{F}(0, 0)$.

At the point $(x, y) = (0, 0)$ the set of linearized feasible directions $\mathcal{F}(0, 0)$ equals

$$\begin{aligned} \mathcal{F}(0, 0) &= \{(p, q) \in \mathbb{R}^2 : (p, q)\nabla c_1(0, 0) \geq 0 \text{ and } (p, q)\nabla c_2(0, 0) \geq 0\} \\ &= \{(p, q) \in \mathbb{R}^2 : (p, q)\begin{pmatrix} 0 \\ 0 \end{pmatrix} \geq 0 \text{ and } (p, q)\begin{pmatrix} 1 \\ -2 \end{pmatrix} \geq 0\} \\ &= \{(p, q) \in \mathbb{R}^2 : p \geq 2q\}. \end{aligned}$$

In particular, the vector $(1,-1)$ is an element of $\mathcal{F}(0,0)$. We show now that $(1,-1)$ cannot be an element of the tangent cone to Ω at $(0,0)$. Indeed, assume to the contrary that $(1,-1)$ is an element of this tangent cone. Then there exists a sequence $(x_k, y_k) \to 0$ and positive scalars t_k converging to 0 such that $(x_k, y_k) \in \Omega$ for all k, $x_k/t_k \to 1$, and $y_k/t_k \to -1$. In particular, this implies that $x_k > 0$ for sufficiently large k whereas $y_k < 0$ for sufficiently large k. This, however, implies that $x_k y_k < 0$, which is a contradiction to the requirement that $(x_k, y_k) \in \Omega$. Thus $(1,-1)$ cannot be an element of the tangent cone at 0.

c) Use the second order optimality conditions in order show that the point $(0,0)$ is a local solution of the constrained optimization problem.

We first have to show $(0,0)$ is a KKT point. That is, we have to find $\lambda \geq 0$ and $\mu \geq 0$ such that $\nabla\mathcal{L}(0,0;\lambda,\mu) = 0$ (note that the point $(0,0)$ is an element of Ω and that both constraints are active; thus we need not care about the other KKT-conditions). The Lagrangian in this case reads as

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

Thus

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

For $(x,y) = (0,0)$ we have

$$\nabla\mathcal{L}(0,0;\lambda,\mu) = \begin{pmatrix} 2-\mu \\ -4+2\mu \end{pmatrix},$$

and we see that $(0,0)$ is a KKT point with Lagrange parameters $\mu = 2$ and arbitrary $\lambda \geq 0$.

Next we note that the Hessian of the Lagrangian with respect to x and y is

$$\nabla^2\mathcal{L}(x,y;\lambda,\mu) = \begin{pmatrix} 2 & -\lambda \\ -\lambda & 2\mu \end{pmatrix}.$$

With the Lagrange parameters $\lambda = 0$ and $\mu = 2$ we see that the Hessian of the Lagrangian at $(0,0)$ is positive definite. Thus the second order sufficient conditions for a strict local solution of a constraint optimization problem are satisfied at the point $(0,0)$.

Problem 4 A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called *strongly convex*, if there exists $c > 0$ such that the function $x \mapsto f(x) - \frac{c}{2}\|x\|^2$ is convex.
A function $F\colon \mathbb{R}^\ell \times \mathbb{R}^m \to \mathbb{R}$ is called *strongly separately convex*, if there exists $c > 0$ such that for every $\hat{x} \in \mathbb{R}^\ell$ and $\hat{y} \in \mathbb{R}^m$ the functions

$$x \mapsto F(x, \hat{y}) - \frac{c}{2}\|x\|^2 \qquad \text{and} \qquad y \mapsto F(\hat{x}, y) - \frac{c}{2}\|y\|^2$$

are convex.

a) Show that a twice differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ is strongly convex, if and only if there exists $c > 0$ such that

$$p^T \nabla^2 f(x) p \ge c\|p\|^2$$

for every $p \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$.

A twice differentiable function g is convex, if and only if $\nabla^2 g(x)$ is positive semi-definite for all x, which is equivalent to $p^T \nabla^2 g(x) p \ge 0$ for all x and p. Applying this characterization to the function $g(x) = f(x) - \frac{c}{2}\|x\|^2$ with Hessian $\nabla^2 g(x) = \nabla^2 f(x) - cI$ with $I \in \mathbb{R}^{n \times n}$ being the identity matrix, we obtain that f is strongly convex, if and only if $\nabla^2 f(x) - cI$ is positive semi-definite for some $c > 0$ and all $x \in \mathbb{R}^n$, which in turn is equivalent to the condition that

$$p^T(\nabla^2 f(x) - cI)p \ge 0$$

for all $x \in \mathbb{R}^n$ and $p \in \mathbb{R}^n$, or, equivalently,

$$p^T \nabla^2 f(x) p \ge c\|p\|^2$$

for all $x \in \mathbb{R}^n$ and $p \in \mathbb{R}^n$.

b) Show that the function $F\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

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

is non-convex, but strongly separately convex.

We note that $\partial_{xx} F(x, y) = 12x^2 + 2 \ge 2$ and $\partial_{yy} F(x, y) = 12y^2 + 2 \ge 2$, showing the F is strongly separately convex. On the other hand,

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

which is not positive semi-definite for $(x, y) = (0, 0)$, as $\det \nabla^2 F(0, 0) = -12 < 0$. Thus F is non-convex.

c) Assume that the function $f\colon \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and strongly convex and assume that x^* is a minimizer of f. Show that there exists $c > 0$ such that

$$f(x) \ge f(x^*) + \frac{c}{2}\|x - x^*\|^2$$

for every $x \in \mathbb{R}^n$.

An application of Taylor's theorem yields

$$f(x) = f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)\nabla^2 f(x + t(x - x^*))(x - x^*)$$

for some $0 \le t \le 1$. Since x^* minimizes f, we have $\nabla f(x^*) = 0$. Moreover, the application of part **a)** of this problem allows us to estimate

$$(x - x^*)\nabla^2 f(x + t(x - x^*))(x - x^*) \ge c\|x - x^*\|^2.$$

Therefore

$$f(x) \ge f(x^*) + \frac{c}{2}\|x - x^*\|^2.$$

d) Assume that the function $F\colon \mathbb{R}^\ell \times \mathbb{R}^m \to \mathbb{R}$ is strongly separately convex, twice continuously differentiable, and coercive. Given $y_0 \in \mathbb{R}^\ell$ we define iterates $x_{k+1} \in \mathbb{R}^\ell$, $y_{k+1} \in \mathbb{R}^m$ by

$$\begin{aligned} x_{k+1} &\text{ minimizes the function } x \mapsto F(x, y_k), \\ y_{k+1} &\text{ minimizes the function } y \mapsto F(x_{k+1}, y). \end{aligned}$$

Show that

$$\sum_{k\in\mathbb{N}} \|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 < \infty.$$

Let $c > 0$ be such that the conditions defining strong separate convexity of F are satisfied. By construction, the point x_{k+1} minimizes the function $x \mapsto F(x, y_k)$. Thus part **c)** implies that

$$F(x_k, y_k) \ge F(x_{k+1}, y_k) + \frac{c}{2}\|x_{k+1} - x_k\|^2$$

or

$$\|x_{k+1} - x_k\|^2 \le \frac{2}{c}(F(x_k, y_k) - F(x_{k+1}, y_k)).$$

Similarly, y_{k+1} minimizes the function $y \mapsto F(x_{k+1}, y)$, and therefore

$$F(x_{k+1}, y_k) \ge F(x_{k+1}, y_{k+1}) + \frac{c}{2}\|y_{k+1} - y_k\|^2$$

or

$$\|y_{k+1} - y_k\|^2 \leq \frac{2}{c}\big(F(x_{k+1}, y_k) - F(x_{k+1}, y_{k+1})\big).$$

Thus

$$\begin{aligned}
\sum_{k\in\mathbb{N}} &\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 \\
&\leq \frac{2}{c}\sum_{k\in\mathbb{N}} F(x_k, y_k) - F(x_{k+1}, y_k) + F(x_{k+1}, y_k) - F(x_{k+1}, y_{k+1}) \\
&= \frac{2}{c}\sum_{k\in\mathbb{N}} F(x_k, y_k) - F(x_{k+1}, y_{k+1}) \\
&= \frac{2}{c}(F(x_1, y_1) - \lim_k F(x_k, y_k) \\
&\leq \frac{2}{c}\big(F(x_1, y_1) - \inf F(x, y)\big).
\end{aligned}$$

Since F is coercive and twice differentiable, it follows that F attains a minimum, and therefore $\inf F(x,y) > -\infty$. Thus

$$\sum_{k\in\mathbb{N}} \|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 < \infty.$$

e) Show that every strongly convex and differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ has a unique minimizer $x^* \in \mathbb{R}^n$.

Let $c > 0$ be such that $g(x) := f(x) - \frac{c}{2}\|x\|^2$ is convex. Then the differential characterization of convexity implies that

$$f(x) - \frac{c}{2}\|x\|^2 = g(x) \geq g(0) + \nabla g(0)^T x = f(0) + \nabla f(0)^T x$$

and therefore

$$f(x) \geq f(0) + \nabla f(0)^T x + \frac{c}{2}\|x\|^2 \geq f(0) - \|\nabla f(0)\|\|x\| + \frac{c}{2}\|x\|^2.$$

In particular, if $\|x\| \to \infty$, then also $f(x) \to \infty$; that is, the function f is coercive. Since f by assumption is differentiable and therefore continuous, it follows that f attains its minimum.

Next assume that x and y are minimizers of f. From **c)** of this problem we obtain that

$$f(x) \geq f(y) + \frac{c}{2}\|x - y\|^2.$$

However, $f(x) = f(y)$ because by assumption both x and y are minimizers of f. Thus

$$\|x - y\|^2 \leq 0,$$

which is only possible if $x = y$. This shows that the minimizer is unique.