**NTNU**
Norwegian University of
Science and Technology

Department of Mathematical Sciences

# Examination paper for **TMA4180 Optimization I**

**Academic contact during examination:** Markus Grasmair
**Phone:** +47 975 80 435

**Examination date:** 08 june 2018
**Examination time (from–to):** 09:00–13:00
**Permitted examination support material:**

- The textbook: Nocedal & Wright, Numerical Optimization including errata
- K. Rottmann: Mathematical formulae
- Handouts on "Minimisers of Optimisation Problems", "Basic Properties of Convex Functions", "Convergence of descent methods with backtracking (Armijo) linesearch. Bisection algorithm for weak Wolfe conditions", "Introduction to optimality conditions: Optimality conditions for optimization over convex sets", "Representation theorem for polyhedral sets"
- 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 the questions of the exam either in English or Norwegian.

**Language:** English
**Number of pages:** 9
**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 $f : \mathbb{R}^2 \to \mathbb{R}$ given by

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

**a)** Determine whether the function $f$ is convex or not.
*(5 points)*

*Solution*: For, example, one can argue like this. A twice differentiable function is convex iff its Hessian matrix is positive semidefinite. For our function, however, we have:

$$\nabla^2 f(0, 0) = \nabla^2[-x^2 + 2x + 2]|_{(x,y)=(0,0)} = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix},$$

where we ignore all polynomial terms of degree 3 or higher since their second derivatives are 0 at $(0, 0)$. Thus $\nabla^2 f(0, 0)$ is negative semidefinite (the eigenvalues of a diagonal matrix are elements on the diagonal), and the function is not convex.

**b)** Consider the point $(\hat{x}, \hat{y}) = (0, 0)$. Let $p$ be the steepest descent direction at $(\hat{x}, \hat{y})$. For which values of the parameters $c_1 \in (0, 1)$ and $c_2 \in (0, 1)$ is the step of length $\alpha = 1/2$ from $(\hat{x}, \hat{y})$ along $p$ acceptable for a linesearch routine based on Wolfe conditions?
*(10 points)*

*Solution*: First of all, we compute $p$:

$$p = -\nabla f(\hat{x}, \hat{y}) = -\begin{bmatrix} 4\hat{x}^3\hat{y}^2 + 4\hat{x}^3 - 6\hat{x}^2\hat{y} - 4\hat{x}\hat{y} - 2\hat{x} + 2 \\ 2\hat{x}^4\hat{y} - 2\hat{x}^3 - 2\hat{x}^2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}.$$

The step $\alpha = 1/2$ is acceptable with respect to the sufficient decrease condition, if $f(\hat{x} + \alpha p_1, \hat{x} + \alpha p_2) \leq f(\hat{x}, \hat{y}) + c_1\alpha\nabla f(\hat{x}, \hat{y})^{\mathrm{T}}p$, or $f(-1, 0) \leq 2 - 2c_1$, or $c_1 \leq 1$.

Similarly, the step is acceptable with respect to the curvatire condition provided that $\nabla f(\hat{x}+\alpha p_1, \hat{x}+\alpha p_2)^{\mathrm{T}}p \geq c_2\nabla f(\hat{x}, \hat{y})^{\mathrm{T}}p$. Substituting the numbers we get $[0, 0][2, 0]^{\mathrm{T}} \geq -4c_2$, or $c_2 \geq 0$.

**c)** Compute all stationary points of $f$ and find all local minimizers of $f$.
*(15 points)*

*Solution*: Critical points of $f$ are the solutions of the system:

$$\nabla f(x,y) = \begin{bmatrix} 4x^3y^2 + 4x^3 - 6x^2y - 4xy - 2x + 2 \\ 2x^4y - 2x^3 - 2x^2 \end{bmatrix}$$
$$= \begin{bmatrix} 4x^3y^2 + 4x^3 - 6x^2y - 4xy - 2x + 2 \\ x^2(2x^2y - 2x - 2) \end{bmatrix} = 0.$$

The second equation implies either $x = 0$ or $2x^2y - 2x - 2 = 0$. In the former case, the first equation simplifies to $2 = 0$, which is impossible, therefore the latter alternative must be true: $2(x^2y - x - 1) = 0$. As a result, in the first equation we can simplify certain terms, for example:

$$\begin{aligned} &4x^3y^2 + 4x^3 - 6x^2y - 4xy - 2x + 2 \\ &= 4xy(\underbrace{x^2y - x - 1}_{=0}) + 4x^3 - \underbrace{2x^2y}_{=2x+2} - 2x + 2 \\ &= 4x^3 - 4x = 0, \end{aligned}$$

which has solutions $x_1 = 0$, $x_2 = 1$, and $x_3 = -1$. Going back to the second equation $x^2y - x - 1 = 0$, we see that the first alternative is impossible, whereas for the other two we get $y_2 = 2$ and $y_3 = 0$.

In order to determine the type of these critical points we compute the Hessian:

$$\nabla^2 f(x,y) = \begin{bmatrix} 12x^2y^2 + 12x^2 - 12xy - 4y - 2 & 8x^3y - 6x^2 - 4x \\ 8x^3y - 6x^2 - 4x & 2x^4 \end{bmatrix}$$

At $(x_2, y_2)$ this evaluates to

$$\nabla^2 f(1,2) = \begin{bmatrix} 26 & 6 \\ 6 & 2 \end{bmatrix},$$

which is positive definite (both leading principal minors are positive). Similarly, at $(x_3, y_3)$ we get

$$\nabla^2 f(-1,0) = \begin{bmatrix} 10 & -2 \\ -2 & 2 \end{bmatrix},$$

which is also positive definite (both leading principal minors are positive).

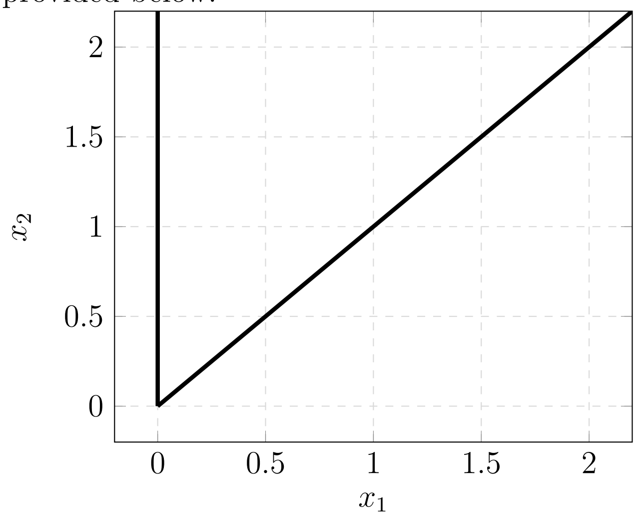
Therefore, both critical points are points of local minima.[1]

[1] In fact, both of them are also global minima, since these are the only points where the non-negative function $f$ equals 0.

**Problem 2** Consider the constrained optimization problem in two variables $(x_1, x_2) \in \mathbb{R}^2$:

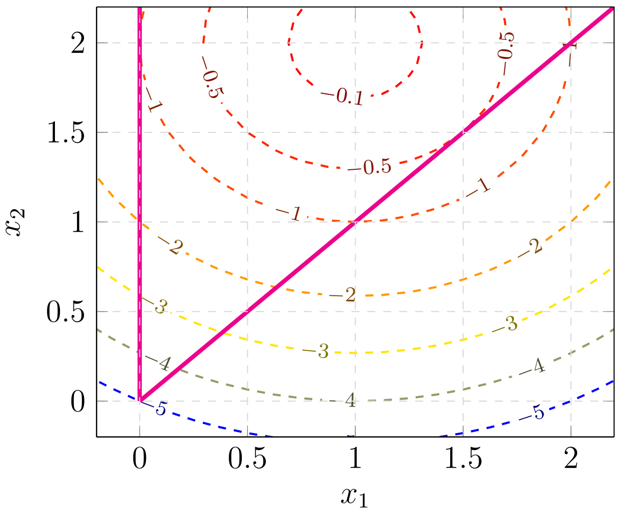
$$\begin{aligned} \text{minimize } & f(x_1, x_2) = -(x_1 - 1)^2 - (x_2 - 2)^2, \\ \text{subject to } & x_1 \geq 0, \\ & x_2 \geq x_1, \\ & x_1 x_2 = x_1^2. \end{aligned} \tag{1}$$

We will denote the feasible set of this problem with $\Omega$. The sketch of this set is provided below:




**a)** Find all globally or locally optimal solutions to (1).
*(5 points)*

*Solution*: Here is the sketch of the feasible set (thick solid lines) and the contour plot of the objective function for future reference:




One can see that the problem has one local minimum at $(0, 0)$, and the function is unbounded from below on the feasible set - so no global minima.

Other points of interest could be the local maxima $(0, 2)$ and $(1.5, 1.5)$, where the contourlines of the objective are tangential to the feasible set; they will play a role in **d)**.

**b)** Demonstrate that LICQ fails at each point in $\Omega$.
*(5 points)*

*Solution*: At $(x_1, x_2) = (0, 0)$ all three constraints are active, and therefore their gradients cannot be linearly independent in 2D space.

On the ray $\{ (0, x_2) : x_2 > 0 \}$ we have constraints $x_1 \geq 0$ and $x_1(x_2 - x_1) = 0$ active. Their gradients on this ray are $[1, 0]^{\mathrm{T}}$ and $[x_2 - 2x_1, x_1]^{\mathrm{T}} = [x_2, 0]^{\mathrm{T}}$, clearly linearly dependent.

Sinilarly, on the ray $\{ (x_1, x_1) : x_1 > 0 \}$ the constraints $x_2 - x_1 \geq 0$ and $x_1(x_2 - x_1) = 0$ are active. Their gradients on this ray are $[-1, 1]^{\mathrm{T}}$ and $[x_2 - 2x_1, x_1]^{\mathrm{T}} = [-x_1, x_1]^{\mathrm{T}}$, clearly linearly dependent.

Thus LICQ fails everywhere in the domain.

**c)** Identify all points $x \in \Omega$ where the strict inclusion $\mathcal{T}_\Omega(x) \subsetneq \mathcal{F}_\Omega(x)$ holds, where $\mathcal{T}_\Omega(x)$ is the tangent cone and $\mathcal{F}_\Omega(x)$ is the cone of linearized feasible directions for $\Omega$ at $x$.
*(10 points)*

*Solution*: On the ray $\{ (0, x_2) : x_2 > 0 \}$ the cone of linearized feasible directions is $\mathcal{F}_\Omega(0, x_2) = \{ [p_1, p_2]^{\mathrm{T}} \mid [1, 0][p_1, p_2]^{\mathrm{T}} \geq 0 \wedge [x_2, 0][p_1, p_2]^{\mathrm{T}} =$

$0\,\} = \{\, [0, p_2]^{\mathrm{T}} \mid p_2 \in \mathbb{R}\,\}$. Locally around points on this ray the feasible set is a straight line $x_1 = 0$, therefore the tangent set is also $\mathcal{T}_\Omega(0, x_2) = \{\, [0, p_2]^{\mathrm{T}} \mid p_2 \in \mathbb{R}\,\} = \mathcal{F}_\Omega(0, x_2)$.

An absolutely similar situation happens on the ray $\{\, (x_1, x_1) : x_1 > 0\,\}$.

At the point $(x_1, x_2) = 0$ we compute the cone of linearized feasible directions as $\mathcal{F}_\Omega(0,0) = \{\, [p_1,p_2]^{\mathrm{T}} \mid [1,0][p_1,p_2]^{\mathrm{T}} \geq 0 \wedge [-1,1][p_1,p_2]^{\mathrm{T}} \geq 0 \wedge [0,0][p_1,p_2]^{\mathrm{T}} = 0\,\} = \{\, [p_1,p_2]^{\mathrm{T}} \mid p_2 \geq p_1 \geq 0\,\}$. However the feasible set is just a union of two straight rays, therefore the tangent cone at this point will be $\mathcal{T}_\Omega(0,0) = \{\, [p_1,p_2]^{\mathrm{T}} \mid p_2 \geq 0, p_1 = 0\,\} \cup \{\, [p_1,p_2]^{\mathrm{T}} \mid p_2 = p_1 \geq 0\,\} = \Omega \subsetneq \mathcal{F}_\Omega(0,0)$.

**d)** Identify all points satisfying the KKT conditions for this problem and compute the corresponding Lagrange multipliers.
*(10 points)*

*Solution*: Graphically on the ray $\{\, (0, x_2) : x_2 > 0\,\}$ there is only one point where the gradients of the active constraints are parallel with the gradient of the objective function, namely the point $(0, 2)$. The Lagrange multipliers, if exist, must satisfy the equations

$$\begin{aligned} \nabla f(0,2) = [2,0]^{\mathrm{T}} &= \lambda_1[1,0]^{\mathrm{T}} + \lambda_3[2,0]^{\mathrm{T}}, \\ \lambda_2 &= 0, \\ \lambda_1 &\geq 0, \end{aligned}$$

where $\lambda_2 = 0$ because the second constraint is inactive. Thus the Lagrange multipliers can be parametrized as $[\lambda_1, 0, 1 - \lambda_1/2]^{\mathrm{T}}$, $\lambda_1 \geq 0$.

A similar calculation on the ray $\{\, (x_1, x_1) : x_1 > 0\,\}$ at the point $(1.5, 1.5)$ gives the system

$$\begin{aligned} \nabla f(1.5,1.5) = [-1,1]^{\mathrm{T}} &= \lambda_2[-1,1]^{\mathrm{T}} + \lambda_3[-1.5,1.5]^{\mathrm{T}}, \\ \lambda_1 &= 0, \\ \lambda_2 &\geq 0, \end{aligned}$$

and the final answer $[0, \lambda_2, \frac{2}{3}(1 - \lambda_2)]^{\mathrm{T}}$, $\lambda_2 \geq 0$.

At the point of local minimum, $(0, 0)$, we have the system

$$\begin{aligned} \nabla f(0,0) = [2,4]^{\mathrm{T}} &= \lambda_1[1,0]^{\mathrm{T}} + \lambda_2[-1,1]^{\mathrm{T}} + \lambda_3[0,0]^{\mathrm{T}}, \\ \lambda_1 &\geq 0, \\ \lambda_2 &\geq 0. \end{aligned}$$

This is solvable by $[6, 4, \lambda_3]^{\mathrm{T}}$, $\lambda_3 \in \mathbb{R}$.

**Problem 3** Consider the linear programming problem

$$\begin{aligned} &\text{minimize } -x_2 + x_3, \\ &\text{subject to } \quad -2x_2 + \tfrac{1}{2}x_3 = 1, \\ &\qquad -x_1 - x_2 + x_3 = 1, \\ &\qquad x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{aligned}$$

**a)** Formulate the dual of this problem and solve it (you can solve it graphically). *(5 points)*

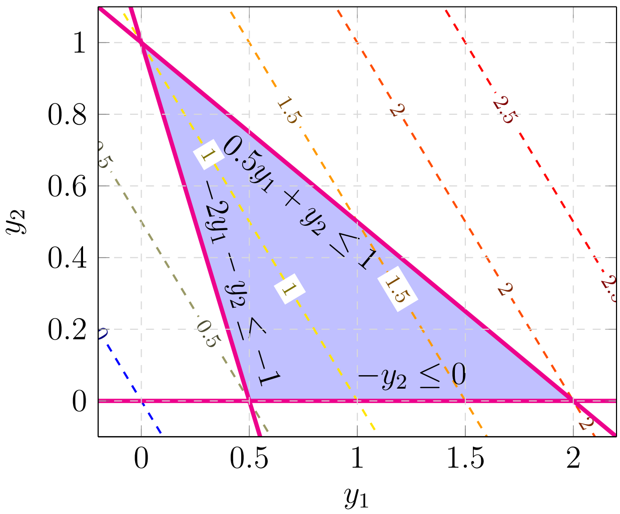
*Solution*: Let us define

$$A = \begin{bmatrix} 0 & -2 & 1/2 \\ -1 & -1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad c = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

The dual problem can be compactly written as

$$\begin{aligned} &\text{maximize } b^{\mathrm{T}}y, \\ &\text{subject to } A^{\mathrm{T}}y \leq c. \end{aligned}$$

The feasible set of the dual problem including the level sets of the objective function are shown in the figure below:




Evidently the optimal solution is $(y_1^*, y_2^*) = (2, 0)$.

**b)** Using **a)**, recover the optimal solution to the primal problem. *(10 points)*

*Solution*: Note that constraint #2 in the figure above is inactive at the optimal solution, which implies (from the complementarity conditions $(x^*)^{\mathrm{T}}(c - A^{\mathrm{T}}y^*) = 0$) that $x_2^* = 0$. Weak duality implies that $c^{\mathrm{T}}x^* = -x_2^* + x_3^* = x_3^* = b^{\mathrm{T}}y^* = 2$, and therefore $x_3^* = 2$. At last, $x_1^*$ can be found from the equality constraint $-x_1^* - x_2^* + x_3^* = 1$, yeilding $x_1^* = 1$.

**Problem 4** Consider the following (albeit very artificial) constrained optimization problem:

$$\begin{aligned} &\text{minimize } (x-b)^2, \\ &\text{subject to } x = 0, \end{aligned}$$

where $x \in \mathbb{R}$ is the optimization variable and $b \in \mathbb{R}$ is a given number.

**a)** State and solve the KKT system corresponding to this optimization problem. *(5 points)*

*Solution*: Just for the record, we note that KKT conditions are both necessary (constraints are linear $\implies$ CQ holds) and sufficient (convex problem) for optimality.

KKT system:

$$\begin{cases} 2(x-b) = \lambda \\ x = 0 \end{cases} \iff \begin{cases} x = 0 \\ \lambda = -2b \end{cases}$$

**b)** Consider the exact penalty approximation of our problem:

$$\text{minimize}_{x\in\mathbb{R}}(x-b)^2 + \mu|x|,$$

where $\mu \geq 0$ is a penalty parameter. The function $S_\mu(b) = \operatorname{argmin}_{x\in\mathbb{R}}[(x-b)^2 + \mu|x|]$ is known as *soft thresholding* in machine learning. Express the solution to the exact penalty problem as an explicit function of $b \in \mathbb{R}$ and $\mu \geq 0$ and sketch the graph of $S_1(\cdot)$.
*(15 points)*

*Solution*: The objective function of the penalty problem is convex as a sum of two convex functions, as well as continuous and coercive, so the optimal solution exists for all $\mu \geq 0$. It is also differentiable everywhere except the point $x = 0$, where it is only directionally differentiable. If we can find a point $x^* \neq 0$ where $f'_\mu(x^*) = ((x^*-b)^2 + \mu|x^*|)' = 0$ then $x^*$ must be a local minimum in some convex neighbourhood where the function is differentiable (first order optimality conditions are sufficient under convexity), and therefore also a global minimum over the whole real line. If the derivative is never zero for $x \neq 0$, then $x^* = 0$ must be the optimal solution.

Let us execute this strategy. For $x \neq 0$ we compute

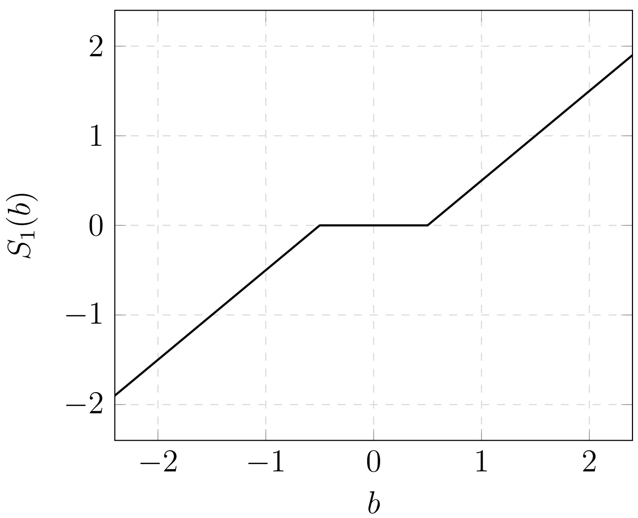
$$f'_\mu(x) = 2(x-b) + \mu \operatorname{sign}(x) = 0 \quad \Longleftrightarrow \quad x = b - \tfrac{\mu}{2}\operatorname{sign}(x).$$

The last equality is solvable for $x \neq 0$ when $|b| > \mu/2$, in which case $x^* = b - \frac{\mu}{2}\operatorname{sign}(b)$. Otherwise the only remaining possibility for $x^*$ is $x^* = 0$. (Note that when $\mu$ is bigger than the absolute value of the Lagrange multiplier, i.e. if $\mu > |2b|$, corresponding to the optimal solution for our constrained problem, we recover its exact solution, i.e. $x^* = 0$.)

We can now explicitly define $S_\mu(b)$:

$$S_\mu(b) = \begin{cases} b - \mu/2, & b > \mu/2, \\ 0, & |b| \leq \mu/2, \\ b + \mu/2, & b < -\mu/2. \end{cases}$$

The plot of $S_1(b)$ is given below:




**Problem 5** Consider the optimization problems

$$\begin{aligned} &\text{minimize } f(x), \\ &\text{subject to } x \in \mathbb{R}^n : x^{\mathrm{T}} x \leq 1, \end{aligned} \tag{2}$$

and

$$\begin{aligned} &\text{minimize } f(x), \\ &\text{subject to } x \in \mathbb{R}^n : x^{\mathrm{T}} x = 1, \end{aligned} \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous, not necessarily differentiable, *concave* function. The feasible set of each of these problems is non-empty, closed, and bounded, and therefore either problem admits globally optimal solutions (you do not need to prove this).

Show that each globally optimal solution to (3) is also a globally optimal solution to (2).
*(15 points)*

*Solution*: Let us denote the feasible set of (2) with $\Omega_1$, and that of (3) with $\Omega_2$. Note that $\Omega_1 \supsetneq \Omega_2$, and therefore $\min_{x\in\Omega_1} f(x) \leq \min_{x\in\Omega_2} f(x)$.

On the other hand, for each $x \in \Omega_1$ we can represent it as a convex combination of (two) points from $\Omega_2$. Indeed, if $x = (0, 0)$ then for example $(0, 0) = \frac{1}{2}(1, 0) + \frac{1}{2}(-1, 0)$, where $(\pm 1, 0) \in \Omega_2$. If, on the other hand, $1 \geq \|x\|_2 \neq 0$ then we can define $x_+ = x/\|x\|_2 \in \Omega_2$ and $x_- = -x/\|x\|_2 \in \Omega_2$. Then $x = \lambda x_+ + (1-\lambda)x_-$, where $\lambda = (1 + \|x\|_2)/2 \in (1/2, 1]$. Since $f$ is concave we have the inequality $f(x) \geq \lambda f(x_+)+(1-\lambda)f(x_-)$ which can be trivially continued to $f(x) \geq \lambda f(x_+)+(1-\lambda)f(x_-) \geq \lambda \min_{\tilde{x}\in\Omega_2} f(\tilde{x})+(1-\lambda) \min_{\tilde{x}\in\Omega_2} f(\tilde{x}) = \min_{\tilde{x}\in\Omega_2} f(\tilde{x})$, since both $\lambda$ and $1-\lambda$ are non-negative. This inequality holds even if we take $x$ to be a global minimum of (2), thereby yielding the inequality $\min_{x\in\Omega_1} f(x) \geq \min_{x\in\Omega_2} f(x)$.

Thus we have shown that $\min_{x\in\Omega_1} f(x) = \min_{x\in\Omega_2} f(x)$, which given the fact that $\Omega_2 \subsetneq \Omega_1$ is sufficient to conclude that each globally optimal solution to (3) is also a globally optimal solution to (2).