Chalmers/GU
Mathematics sciences

EXAM

TMA947/MMG621
NONLINEAR OPTIMISATION

Date:	24–10–31
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)

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

(Linear programming)

(2p) a) Consider the linear programming problem

$$\begin{aligned} \text{minimize} \quad & x_1 + 2x_2 - x_3 + 3x_4 \\ \text{subject to} \quad & x_1 \qquad\; - 2x_3 + 4x_4 \leq 1 \\ & x_1 - \; x_2 + 2x_3 \qquad\;\; \leq 1 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

Solve the problem using the Simplex method. Start with x_1 and x_3 as basic variables.

Hint: You may find the following identity useful:

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

(1p) b) If you find that 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 ray of unboundedness of the objective value.

(3p) Question 2

(Unconstrained optimization)

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x) = -\frac{1}{12}(x_1 + 1)^3 + x_2^2 - x_2 x_3 + x_3^2$. In the point $\bar{x} = [1, 1, 1]^T$, compute the search direction obtained by the following three unconstrained optimization methods:

a) The steepest descent method,

b) Newton's method,

c) Newton's method with the Levenberg-Marquardt modification. Use a modification parameter $\gamma \geq 0$ which is the smallest integer such that the conditions needed are fulfilled.

Which of the above search directions are descent directions?

Question 3

(Weak duality)

Given a primal problem

$$\begin{cases} \text{minimize} & f(x), \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & x \in X, \end{cases}$$

where $X \subseteq \mathbb{R}^n$ is some base set, we can define a Lagrangian dual function $q(\mu)$ by making a Lagrangian relaxation of the inequality constraints. We can also define a (Lagrangian) dual optimization problem. The weak duality theorem states a relation between the primal and the dual objective function, and a relation between the optimal values f^* (of the primal optimization problem) and q^* (of the dual optimization problem).

(1p) a) Define the Lagrangian, the dual function, and the dual optimization problem. Then give the formal statement of the weak duality theorem.

(2p) b) Prove the weak duality 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 4

(Nonlinear optimization)

For $z \in \mathbb{R}$, define $(z)_+ := \max\{0, z\}$, that is $(z)_+ = 0$ if $z < 0$ and $(z)_+ = z$ if $z \geq 0$. Consider the optimization problem

$$\begin{cases} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\|x - y\|^2, \\ \text{subject to} & \sum_{i=1}^n x_i = 1, \\ & x_i \geq 0, \quad i = 1, \dots, n, \end{cases}$$

where $y \in \mathbb{R}^n$ is some given vector and $\|\cdot\|$ denotes the standard Euclidean norm of a vector. Use results and methods from the course to show that the globally optimal solution to this problem is given by the vector x^* with components $x_i^* = (y_i - \lambda)_+$ for λ such that

$$\sum_{i=1}^n (y_i - \lambda)_+ = 1.$$

Question 5

(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 $f : \mathbb{R}^n \to \mathbb{R}$ be convex and consider the problem

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

Claim: A globally optimal point x^* always exists for this type of problem.

(1p) b) Let $f : \mathbb{R}^3 \to \mathbb{R}$, and let $g_i : \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, 2, 3, 4$, and consider the problem

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, 2, 3, 4. \end{cases}$$

Claim: There exist convex functions f and g_i, for $i = 1, 2, 3, 4$, such that the following is a globally optimal solution to the problem that satisfies the global optimality conditions

$$x^* = \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}, \quad \mu^* = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad f(x^*) = 0, \quad \begin{bmatrix} g_1(x^*) \\ g_2(x^*) \\ g_3(x^*) \\ g_4(x^*) \end{bmatrix} = \begin{bmatrix} -7 \\ 0 \\ 0 \\ -3 \end{bmatrix}.$$

(1p) c) Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = x^T Q x + c^T x$, where $Q \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $c \in \mathbb{R}^n$.

Claim: f is a convex function if Q has full rank.

Question 6

(the Karush-Kuhn-Tucker conditions)

Consider the problem

$$\begin{cases} \text{minimize} & f(x), \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, 2, 3, 4 \\ & x \in \mathbb{R}^2, \end{cases}$$

where

$$\begin{aligned} f(x) &= -\tfrac{3\sqrt{3}}{8\pi} x_1^2 - \tfrac{1}{6} x_2^2, \\ g_1(x) &= -1 - \cos(x_1) + x_2, \\ g_2(x) &= -x_2 + 1, \\ g_3(x) &= x_1 - \pi, \\ g_4(x) &= -x_1 - \pi. \end{aligned}$$

(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 1: Consider the point $x = [\pi/3,\ 3/2]^T$.

Hint 2: The following table of values for sine and cosine might be helpful

$\theta =$	$-\frac{\pi}{2}$	$-\frac{\pi}{3}$	$-\frac{\pi}{4}$	$-\frac{\pi}{6}$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\cos(\theta) =$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\sin(\theta) =$	-1	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

Hint 3: There is no feasible point such that $g_3(x) = 0$, or such that $g_4(x) = 0$.

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

Question 7

(Exterior penalty method)

Consider the problem

$$\begin{cases} \text{minimize} & f(x), \\ \text{subject to} & h(x) = 0, \\ & x \in \mathbb{R}^2. \end{cases}$$

where

$$f(x) = -x_1^2 + x_2^4,$$
$$h(x) = x_1 + x_2.$$

We will consider the exterior penalty method with penalty function $\psi(s) = s^2$.

(0.5p) a) State the penalty transformed problem that needs to be solved in each iteration of the exterior penalty method.

(1.5p) b) Solve the penalty transformed problem. That is, compute an optimal solution $x^*(\nu)$ as a function of ν. There are multiple solutions, and one such solution is $x^*(\nu) = [0, 0]^T$. You have to compute at least one solution which is not $x^*(\nu) = [0, 0]^T$. Moreover, compute the solution under the assumption that the penalty parameter $\nu \geq 2$.

(1p) c) Finally, compute $\lim_{\nu \to \infty} x^*(\nu)$ for one of the solutions which is not $x^*(\nu) = [0, 0]^T$. State what you know about this limit.

Hint: For each $\nu \geq 2$, there are three optimal solutions to the penalty transformed problem. One of them is $x^*(\nu) = [0, 0]^T$. Do not consider this point. Moreover, the two other points attain the same objective function value in the penalty transformed problem. It is enough to consider one of the two solutions.

Chalmers/GU
Mathematics

EXAM SOLUTION

TMA947/MMG621
NONLINEAR OPTIMISATION

Date: 24–10–31
Examiner: Axel Ringh

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

Question 1

(Linear programming)

(2p) a) Transforming the problem to standard form gives

$$\begin{aligned} \text{minimize} \quad & x_1 + 2x_2 - x_3 + 3x_4 \\ \text{subject to} \quad & x_1 \qquad - 2x_3 + 4x_4 + s_1 \qquad = 1 \\ & x_1 - \ x_2 + 2x_3 \qquad\qquad + s_2 = 1 \\ & x_1, x_2, x_3, x_4, s_1, s_2 \geq 0. \end{aligned}$$

As stated in the problem, we start with x_1 and x_3 as basic variables.

Iteration 1:
With $x_B = [x_1,\, x_3]^T$ and $x_N = [x_2,\, x_4,\, s_1,\, s_2]^T$,

$$B = \begin{bmatrix} 1 & -2 \\ 1 & 2 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 4 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad c_B^T = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad c_N^T = \begin{bmatrix} 2 & 3 & 0 & 0 \end{bmatrix}.$$

$x_B = B^{-1}b = [1,\, 0]^T$, and this is thus a BFS (as expected). The reduced costs are

$$\tilde{c}_N^T = \begin{bmatrix} 2 & 3 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & -1 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 4 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} = [9/4,\, 0,\, -3/4,\, -1/4],$$

and hence $(x_N)_3 = s_1$ enters the basis. $B^{-1}N_3 = 1/4[2,\, -1]^T$. The only positive element is the first one, and thus the minimum ratio test gives that $(x_B)_1 = x_1$ leaves the basis

Iteration 2:
With $x_B = [x_3,\, s_1]^T$ and $x_N = [x_1,\, x_2,\, x_4,\, s_2]^T$,

$$B = \begin{bmatrix} -2 & 1 \\ 2 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 1 & 0 & 4 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}, \quad c_B^T = \begin{bmatrix} -1 & 0 \end{bmatrix}, \quad c_N^T = \begin{bmatrix} 1 & 2 & 3 & 0 \end{bmatrix}.$$

$x_B = B^{-1}b = [1/2,\, 2]^T$. The reduced costs are

$$\tilde{c}_N^T = \begin{bmatrix} 1 & 2 & 3 & 0 \end{bmatrix} - \begin{bmatrix} -1 & 0 \end{bmatrix} \frac{-1}{2} \begin{bmatrix} 0 & -1 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix} = [3/2,\, 3/2,\, 3,\, 1/2].$$

Since the reudced costs are larger than or equal o zero, this BFS is optimal. Thus $x^* = [x_1^*,\, x_2^*,\, x_3^*,\, x_4^*]^T = [0,\, 0,\, 1/2,\, 0]^T$.

(1p) b) Since the reduced costs are strictly positive, the solution is unique.

(3p) Question 2

(Unconstrained optimization)

We have that

$$\nabla f(x) = \begin{bmatrix} -\frac{1}{4}(x_1+1)^2 \\ 2x_2 - x_3 \\ 2x_3 - x_2 \end{bmatrix}, \qquad \nabla^2 f(x) = \begin{bmatrix} -\frac{1}{2}(x_1+1) & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

and therefore

$$\nabla f(\bar{x}) = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \qquad \nabla^2 f(\bar{x}) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

We therefore get that:

a) $p_{SD} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$

b) $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} p_{Newton} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$, which gives $p_{Newton} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$.

c) We need to find the smallest integer $\gamma \geq 0$ such that $\nabla^2 f(\bar{x}) + \gamma I$ is positive definite. By the structure of $f(\bar{x})$, the eigenvalues are given by -1 and the eigenvalues of $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. The latter is positive definite (why?), which means that $\gamma = 2$. Therefore, the search direction is given by $\left(\begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \right) p_{LM} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$, which gives $p_{LM} = \begin{bmatrix} 1 \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$.

For all three search directions, it can be verified that $\nabla f(\bar{x})^T p < 0$ (the inner products are -3, -1 and $-5/3$, respectively), which means that they are all descent directions.

(Since $\nabla^2 f(\bar{x})$ is not positive definite, Newton's method is not recommended to use. In fact, it computes a direction in x_1 which is ascent. But the search direction computed in the point still turns out to be a descent direction.)

Question 3

(Weak duality)

See pages 158–161 in the course book; in particular Theorem 6.5.

(3p) Question 4

(Nonlinear optimization)

The optimization problem

$$\begin{cases} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\|x-y\|^2, \\ \text{subject to} & \sum_{i=1}^n x_i = 1, \\ & x_i \geq 0, \quad i = 1, \dots, n, \end{cases}$$

is convex (why?). To solve the problem, we try to use the global optimality conditions. To this end, we consider the Lagrangian

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\|x-y\|^2 + \lambda \left(\sum_{i=1}^n x_i - 1 \right) = \sum_{i=1}^n \left(\frac{1}{2}(x_i - y_i)^2 + \lambda x_i \right) - \lambda,$$

in which case the global optimality conditions are

$$x^* \in \underset{x \geq 0}{\arg\min}\, \mathcal{L}(x, \lambda^*)$$
$$x^* \geq 0, \quad \sum_{i=1}^n x_i^* = 1.$$

(Why are these the optimality conditions? We have only relaxed an equality constraint and not relaxed any inequality constraints.) The problem $\min_{x \geq 0} \mathcal{L}(x, \lambda)$ decouples coordinate-wise, and for $i = 1, \dots, n$, we are thus interested in solving

$$\min_{x_i \geq 0} \frac{1}{2}(x_i - y_i)^2 + \lambda x_i.$$

Since this is a convex problem, the optimal solution is either where the derivative is equal to 0 or at $x_i = 0$. Equating the derivative to 0 gives

$$0 = x_i - y_i + \lambda \qquad \Longrightarrow \qquad x_i = y_i - \lambda.$$

This means that $x_i^* = y_i - \lambda$ if $y_i - \lambda \geq 0$ and $x_i^* = 0$ if $y_i - \lambda < 0$, or in other words that $x_i^* = (y_i - \lambda)_+$. This point fulfills the global optimality conditions if and only if $\sum_{i=1}^n x_i^* = \sum_{i=1}^n (y_i - \lambda)_+ = 1$.

Alternative solution

We can also look for a KKT point. Since the problem is convex, finding a KKT point implies that the point is globally optimal. The KKT system is given by

$$\nabla f(x) + \sum_{i=1}^{n} \mu_i \nabla g_i(x) + \lambda \nabla h(x) = \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{bmatrix} + \sum_{i=1}^{n} \mu_i(-e_i) + \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\mu_i \geq 0, \quad i = 1, \dots, n$$

$$\mu_i x_i = 0, \quad i = 1, \dots, n,$$

and since we also want feasbile points we need that $x_i \geq 0$, $i = 1, \dots, n$ and that $\sum_{i=1}^{n} x_i = 1$.

Now, note that for $x_i^* = (y_i - \lambda)_+$, it holds that $\lambda \geq y_i - x_i^*$ (since $x_i^* = 0$ if $\lambda \geq y_i$ and otherwise $\lambda = y_i - x_i^*$). Moreover, if $x_i^* = 0$ then we can select $\mu_i = \lambda - y_i \geq 0$ where the inequality follows by the previous inequality. Therefore, if $x_i^* = 0$ then $\mu_i = \lambda - y_i$ satisfies the ith row of the KKT system, and if $x_i^* > 0$ then $x_i^* = y_i - \lambda$ and $\mu_i = 0$ satisfies the ith row KKT system. Finally, $x_i^* = (y_i - \lambda)_+ \geq 0$ so for it to be feasible it must fulfill $\sum_{i=1}^{n}(y_i - \lambda)_+ = 1$.

Another alternative solution (by one of the students on the exam)

For any $\lambda \in \mathbb{R}$, consider the relaxed problem

$$\begin{cases} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\|x - y\|^2 + \lambda \left(\sum_{i=1}^{n} x_i - 1\right), \\ \text{subject to} & x_i \geq 0, \quad i = 1, \dots, n, \end{cases}$$

which is a convex optimization problem (why?). Calling this cost function $F_\lambda(x)$, a point $x^{(\lambda)}$ is optimal to the relaxed problem if and only if $\nabla F_\lambda(x^{(\lambda)})^T(x - x^{(\lambda)}) \geq 0$ for all $x \geq 0$, that is, for all feasible x (Theorem 4.23). We have that

$$\nabla F_\lambda(x) = \begin{bmatrix} x_1 - y_1 + \lambda \\ \vdots \\ x_n - y_n + \lambda \end{bmatrix}.$$

We investigate the proposed point $x_i^{(\lambda)} := (y_i - \lambda)_+$. This gives

$$F_\lambda(x^{(\lambda)})^T(x - x^{(\lambda)}) = \sum_{i=1}^{n} \Big(\big((y_i - \lambda)_+ - y_i + \lambda\big)\big(x_i - (y_i - \lambda)_+\big) \Big)$$

If $y_i - \lambda \geq 0$, then $(y_i - \lambda)_+ = y_i - \lambda$ and therefore the first term in the summation is zero. If $y_i - \lambda < 0$, then $(y_i - \lambda)_+ = 0$ and then $\big((y_i - \lambda)_+ - y_i + \lambda\big)\big(x_i - (y_i - \lambda)_+\big) = -(y_i - \lambda)x_i \geq 0$ for all $x_i \geq 0$. Therefore, $F_\lambda(x^{(\lambda)})^T(x - x^{(\lambda)}) \geq 0$ for all $x \geq 0$ and thus $x_i^{(\lambda)} = (y_i - \lambda)_+$ is the globally optimal solution to the relaxed problem. For λ such that $1 = \sum_{i=1}^{n} x_i^{(\lambda)} = \sum_{i=1}^{n}(y_i - \lambda)_+$, the point is also feasible to the original problem and $F_\lambda(x^{(\lambda)}) = \frac{1}{2}\|x^{(\lambda)} - y\|^2$. By the Relaxation Theorem (Theorem 6.1), it is therefore also optimal to the original problem.

Question 5

(True or False)

(1p) a) False. e^x is a counterexample. x is another.

(1p) b) False. Complementarity does not hold.

(1p) c) False. A counterexample is $f(x) = -x^2$, where $Q = [-1]$ is full rank but f is not convex. (Q must be positive semidefinite for f to be convex.)

Question 6

(the Karush-Kuhn-Tucker conditions)

(1p) a) For feasible points x, i.e., x such that $g_i(x) \leq 0$, $i = 1, 2, 3, 4$, the KKT conditions are

$$\nabla f(x) + \sum_{i=1}^{4} \mu_i \nabla g_i(x) = \begin{bmatrix} -\frac{3\sqrt{3}}{4\pi} x_1 \\ -\frac{1}{3} x_2 \end{bmatrix} + \mu_1 \begin{bmatrix} \sin(x_1) \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \mu_3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mu_4 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\mu_1, \mu_2, \mu_3, \mu_4 \geq 0,$$

$$\mu_1(-1 - \cos(x_1) + x_2) = 0, \quad \mu_2(-x_2 + 1) = 0, \quad \mu_3(x_1 - \pi) = 0, \quad \mu_4(-x_1 - \pi) = 0.$$

(1p) b) Drawing the problem, and using hint 1 to consider $[\pi/3,\ 3/2]^T$, hint 2 to see the symmetry of the problem, and hint 3 to not consider $x_1 = \pm\pi$, we considering the points $[0,\ 2]^T$, $[\pm\pi/3,\ 3/2]^T$, and $[\pm\pi/2,\ 1]^T$.

- For $[0,\ 2]^T$, only constrain 1 is active and we find $\mu_1 = 2/3$.
- For $[\pm\pi/3,\ 3/2]^T$, only constrain 1 is active and we find $\mu_1 = 1/2$.
- For $[\pm\pi/2,\ 1]^T$, constraints 1 and 2 are active and we find $\mu_1 = 3\sqrt{3}/8$ and $\mu_2 = 3\sqrt{3}/8 - 1/3 = (9\sqrt{3} - 8)/24$.

(1p) c) The smallest value among the KKT points is obtained at the points $\hat{x} = [\pm\pi/2,\ 1]^T$. Since the feasible region is closed and bounded, and since f is continuous, by Weierstrass' theorem there exists a globally optimal solution. Since the linear independence CQ holds in all points of the feasible region (verify!) the globally optimal solutions must be KKT points (why?). This means that $\hat{x} = [\pm\pi/2,\ 1]^T$ are globally optimal solutions.

Question 7

(Exterior penalty method)

Consider the problem

$$\begin{cases} \text{minimize} & f(x), \\ \text{subject to} & h(x) = 0, \\ & x \in \mathbb{R}^2. \end{cases}$$

where

$$f(x) = -x_1^2 + x_2^4,$$
$$h(x) = x_1 + x_2.$$

We will consider the exterior penalty method with penalty function $\psi(s) = s^2$.

(0.5p) a) The penalty transformed problem that needs to be solved in each iteration of the exterior penalty method is

$$\min_{x \in \mathbb{R}^2} f(x) + \nu\psi(h(x)) = -x_1^2 + x_2^4 + \nu(x_1 + x_2)^2.$$

(1.5p) b) Let $F_\nu(x) := -x_1^2 + x_2^4 + \nu(x_1 + x_2)^2$ denote the cost function of the penalty transformed problem. An optimal solution to the penalty transformed problem must be obtained where $\nabla F_\nu(x) = 0$ since the problem is unconstrained. This gives the equations

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2x_1 + 2\nu(x_1 + x_2) \\ 4x_2^3 + 2\nu(x_1 + x_2) \end{bmatrix}.$$

From these equations, we get that $-2x_1 = 4x_2^3$, or equivalently that $x_1 = -2x_2^3$. Plugging this into the second equation above gives

$$0 = 4x_2^3 + 2\nu(-2x_2^3 + x_2) = 4(1 - \nu)x_2 \left(x_2^2 - \frac{\nu}{2(\nu - 1)} \right),$$

where we note that $\frac{\nu}{2(\nu-1)} > 0$ for all $\nu \geq 2$ (actually for all $\nu > 1$, but you were asked to take $\nu \geq 2$). The nonzero solutions are given by

$$x_2^*(\nu) = \pm\sqrt{\frac{\nu}{2(\nu - 1)}} \qquad \Longrightarrow \qquad x_1^*(\nu) = \mp 2 \left(\frac{\nu}{2(\nu - 1)} \right)^{\frac{3}{2}}.$$

(1p) c) Plugging the above two points into F_ν we see that they take the same value. For the first of these points, we have that

$$\lim_{\nu \to \infty} \begin{bmatrix} -2\left(\frac{\nu}{2(\nu-1)}\right)^{\frac{3}{2}} \\ \sqrt{\frac{\nu}{2(\nu-1)}} \end{bmatrix} = \begin{bmatrix} -2\left(\frac{1}{2}\right)^{\frac{3}{2}} \\ \sqrt{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} =: \hat{x}.$$

Since all functions are C^1, since $\psi'(s) = 2s \geq 0$ for all $s \geq 0$, since $\hat{x}$ is feasible, and since the LICQ holds in $\hat{x}$, we know that $\hat{x}$ is a KKT point to the original problem (see Theorem 13.4 in course book).