The Chinese University of Hong Kong, Shenzhen
SDS · School of Data Science

Final Exam – Solutions

MAT 3007 – Optimization

Spring Semester 2022

Prof. Andre Milzarek, Prof. Zizhuo Wang, Prof. Haoxiang Yang
May 14th, 2022

Points	
E01	18
E02	22
E03	16
E04	10
E05	18
E06	16
Tot.	100

Please read the following instructions carefully:

- You have **120 minutes** to complete the exam.
  Examination Time: **13:00 to 15:00 pm**.
  The exam consists of **six problems** in total.
- You are (**only**) allowed to bring two self-made sheets of A4 paper (with arbitrary notes on both sides of it) for your personal use in this exam.
  Usage of electronic devices or other tools is not allowed.
- Please abide by the honor codes of CUHK-SZ.
- Please make sure to present your solutions and answers in a comprehensible way and give explanations of your steps and results. Write down all necessary steps when answering the questions.
- Violation of the exam policies will be considered as cheating and reported. Consequences of such violation include zero points for the exam and disciplinary actions.
- Good luck!

Exercise 1 (Convexity of Functions): (*18* points)

Consider the function $f(x,y) := e^x/y$ defined on $x \in \mathbb{R}$ and $y > 0$.

a) Given x, is $f(x,y)$ a convex or a concave function of y? Please explain your answers.

b) Given $y > 0$, is $f(x,y)$ a convex or a concave function of x? Please explain your answers.

c) Is $f(x,y)$ a convex or a concave function of (x,y)? Please explain your answers.

d) Define $g(x,y) = f(\sqrt{x}, y)$ for $0 < x < 1$ and $y > 0$. Given $y > 0$, is $g(x,y)$ a convex or a concave function of x on $0 < x < 1$ (or neither convex nor concave)? Please explain your answers.

Solution:

a) Given x, $f(x,y)$ is a convex function of y because $f_{yy}(x,y) = 2e^x/y^3 > 0$ on its definition region. (4pts)

b) Given y, $f(x,y)$ is a convex function of x because $f_{xx}(x,y) = e^x/y > 0$ on its definition region. (4pts)

c) Consider the Hessian matrix of $f(x,y)$:

$$\nabla^2 f(x,y) = \begin{bmatrix} e^x/y & -e^x/y^2 \\ -e^x/y^2 & 2e^x/y^3 \end{bmatrix}$$

We have $\mathrm{tr}(\nabla^2 f(x,y)) = e^x/y \cdot (1 + 1/y^2) \geq 0$ for all $x \in \mathbb{R}$ and $y > 0$. In addition, it holds that $\det(\nabla^2 f(x,y)) = 2e^{2x}/y^4 - e^{2x}/y^4 = e^{2x}/y^4 \geq 0$ (for all $x \in \mathbb{R}$ and $y > 0$). Hence, the Hessian matrix is positive semidefinite on its definition region and f is convex on $\{(x,y) : x \in \mathbb{R},\ y > 0\}$. (5pts)

d) We consider $g_{xx}(x,y) = \frac{1}{4y} e^{\sqrt{x}} \left(\frac{1}{x} - \frac{1}{x^{3/2}} \right) < 0$ when $0 < x < 1$. Therefore, given $y > 0$, $g(x,y)$ is a concave function of x for $0 < x < 1$. (5pts)

For each part, only answering "convex" or "concave" or "neither" correctly without a correct explanation will get 2 points for that part.

Exercise 2 (KKT Conditions): (*22* points)

Consider the nonlinear optimization problem

$$\min_{x \in \mathbb{R}^2} \ f(x) := (x_1 - 1)^2 - x_2 + x_1 x_2 + \frac{1}{4} x_2^2 \qquad \text{s.t.} \qquad g(x) \leq 0, \tag{1}$$

where the constraint function $g : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $g_1(x) := (x_1 - 1)^2 + x_2 - 2$ and $g_2(x) := x_1 - x_2 - 1$.

A sketch of the feasible set $\Omega := \{x \in \mathbb{R}^2 : g(x) \leq 0\}$ is shown below. (*You can refer to/use this sketch if necessary*).

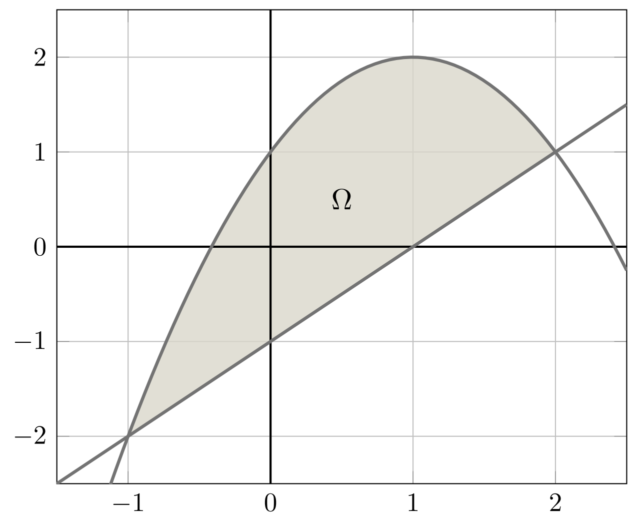
a) Write down the KKT conditions for the optimization problem (1).

b) First consider the following unconstrained version of problem (1):

$$\min_{x\in\mathbb{R}^2} \; f(x) = (x_1-1)^2 - x_2 + x_1x_2 + \frac{1}{4}x_2^2. \quad (2)$$

Find and determine all optimal solutions of the unconstrained problem (2).

c) Find and calculate all KKT points of the original constrained problem (1). Please explain your answer!



Solution:

a) The KKT conditions are given by: find $x \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}^2$ such that:

$$\nabla f(x) + \nabla g_1(x)\lambda_1 + \nabla g_2(x)\lambda_2 = \begin{pmatrix} 2(x_1-1)+x_2 \\ -1+x_1+\frac{1}{2}x_2 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2(x_1-1) \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$\lambda_1, \lambda_2 \geq 0$, $g(x) \leq 0$, and $\lambda_1 g_1(x) = \lambda_2 g_2(x) = 0$.

6 pts in total. 2 pts for correct gradients (-1 for errors; depending whether this leads to strong simplifications). 1 pt for main condition. 1 pt for dual feasibility. 1 pt for primal feasibility. 1 pt for complementarity conditions.

b) We first determine all stationary points of f:

$$\nabla f(x) = 0 \quad\iff x_2 = -2(x_1-1) \text{ and } 0 = -1 + x_1 - (x_1 - 1) = 0.$$

Hence, every point in the set $\mathcal{X} := \{(x, 2(1-x))^\top : x \in \mathbb{R}\}$ is a stationary point of f. We now check the Hessian $\nabla^2 f$ on $\mathcal{X}$:

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 1 \\ 1 & \frac{1}{2} \end{pmatrix}, \quad \mathrm{tr}(\nabla^2 f(x)) = 2.5, \quad \det(\nabla^2 f(x)) = 1 - 1 = 0.$$

This shows that $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^2$ and thus, f is a convex function. Consequently, all points in $\mathcal{X}$ are optimal solutions of the problem (2).

8 pts in total. 3 pts for correct derivation of $\mathcal{X}$. 1 pt for $\nabla^2 f$. 2 pt for checking and verifying positive semidefiniteness of $\nabla^2 f$. 1+1 pt for establishing convexity and global optimality of the stationary points.

c) The function g_2 is linear and we have

$$\nabla^2 g_1(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence, g_1 is a convex function and utilizing the result from part b), we can infer that problem (1) is convex. (Convexity of Ω basically can also be inferred by considering the figure). Hence, every

KKT point of problem (1) is an optimal solution and every optimal solution of (1) has to satisfy the KKT conditions. From the sketch, we see $\Omega \cap \mathcal{X} \neq \emptyset$. Since all points in $\Omega \cap \mathcal{X}$ are global solutions of $\min_x f(x)$, these points will also be global solutions of $\min_{x\in\Omega} f(x)$. Furthermore, we have $x \in \Omega \cap \mathcal{X}$ if and only if

$$x_1 + 2(x_1) - 1 \leq 0 \iff x_1 \geq 1 \quad \text{and} \quad (x_1 - 1)^2 - 2(x_1 - 1) - 2 \leq 0.$$

The last conditions requires $x_1 \geq 2 - \sqrt{3}$ and $x_1 \leq 2 + \sqrt{3}$. Together, this shows that all points in $\mathcal{X}^* = \{(x, 2(1-x))^\top : x \in [2-\sqrt{3}, 1]\}$ are global solutions and KKT points of problem (1). Since KKT points correspond to global solutions, this problem cannot have any other additional KKT points.

8 pts. 2+1 pts for showing convexity of the problem and mentioning that KKT points are global solutions. 5 pts for deriving $\mathcal{X}^*$ (different possibilities; steps: 2 pts for – all points in $\mathcal{X} \cap \Omega$ are global solutions of (1); 2 pts for – calculate $\mathcal{X} \cap \Omega$; 1 pt – no other points can be KKT points).

Exercise 3 (True or False with Explanations): (*16* points)

For each statement, state whether it is true or false. If it is true, then provide appropriate explanations. If it is false, then present a counterexample.

a) Suppose f and g are both twice continuously differentiable functions on $\mathbb{R}$. Suppose f is a convex and decreasing function, g is a concave function. Then, $f(g(x))$ must be a convex function.

b) Consider the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

where f is a convex function. If there are two distinct optimal solutions, then there must be an infinite number of optimal solutions.

c) Consider an integer program (S1) which is a maximization problem. We denote its linear programming relaxation by (S2). Consider the dual of (S2) which we denote by (S3). Then we impose integer constraints on (S3) for all variables and the resulting problem is denoted by (S4). If both (S1) and (S4) are feasible and bounded, then it must be that the optimal value of (S1) is less than or equal to that of (S4).

d) For the following statement, you only need to state "true" or "false", no explanation is needed: For a deep learning problem where a billion parameters need to be optimized (with a nonlinear objective function), it is more suitable to use the gradient descent method than Newton's method.

Solution:

a) **True.** The second-order derivative of $f(g(\cdot))$ is

$$(f(g(x)))'' = f''(g(x))(g'(x))^2 + f'(g(x))g''(x)$$

If f is convex and decreasing, and g is concave, then the second-order derivative is non-negative. Therefore, $f(g(\cdot))$ is a convex function. (4 pts)

b) **False.** Consider the following problem:

$$\begin{array}{rl} \text{maximize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 + x_2 = 1 \\ & x_1, x_2 \geq 0 \end{array}$$

There are two and only two optimal solutions to this problem, which are $(1, 0)$ and $(0, 1)$. (4 pts)

c) **True.** From (S1) to (S2) is because relaxing the integer constraints for a maximization problem will make the solution bigger; from (S2) to (S3) is because of the weak duality theorem; from (S3) to (S4) is because imposing integer constraints to a minimization problem (P3) would make the optimal value larger. (4 pts)

d) **True.** (4 pts)

For part (a)-(c), only answering the "true" or "false" part correctly without a valid explanation will get 2 points for that part.

Exercise 4 (Algorithms for Unconstrained Problems): (*10* points)

We consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^2} \; f(x) := \frac{1}{2}(x_1^2 - 2x_1x_2)^2 - \frac{1}{3}x_1^3 + x_1^2 + 2x_1 - 2x_1x_2 + 2x_2^2. \tag{3}$$

It can be shown that the two points $x^* = (2, 1)^\top$ and $y^* = (-1, -\frac{1}{2})^\top$ are the only stationary points of the problem (3). (*You don't need to verify this!*)

a) Let $x \in \mathbb{R}^2$ be given with $x \neq x^*, y^*$ and let us define the direction $d_\infty = -(\nabla f(x)_j) \cdot e_j = -\frac{\partial f}{\partial x_j}(x) \cdot e_j$, where $e_j \in \mathbb{R}^2$ is the j-th unit-vector and $j \in \{1, 2\}$ is an index satisfying

$$\left| \frac{\partial f}{\partial x_j}(x) \right| = \max \left\{ \left| \frac{\partial f}{\partial x_1}(x) \right|, \left| \frac{\partial f}{\partial x_2}(x) \right| \right\} = \|\nabla f(x)\|_\infty.$$

Show that d_∞ is a descent direction of f at x.

b) We use $(2, 2)^\top$, $(-2, -2)^\top$, and $(-2, 2)^\top$ as initial points and apply the following methods to solve this problem:

1. The gradient method with backtracking.
2. The gradient method with exact line search.
3. The globalized Newton method with backtracking.
4. A gradient descent-type method using the directions d_∞ from part a) as descent direction and backtracking.

The solution paths of the different methods are shown in Figure 1 below. Match each of the subfigures in Figure 1 to one of the algorithms, i.e., give a list of matching pairs "[`subfigure`, `algorithm`]" (e.g., $[(a), 1]$). Each of the algorithms is only applied once!

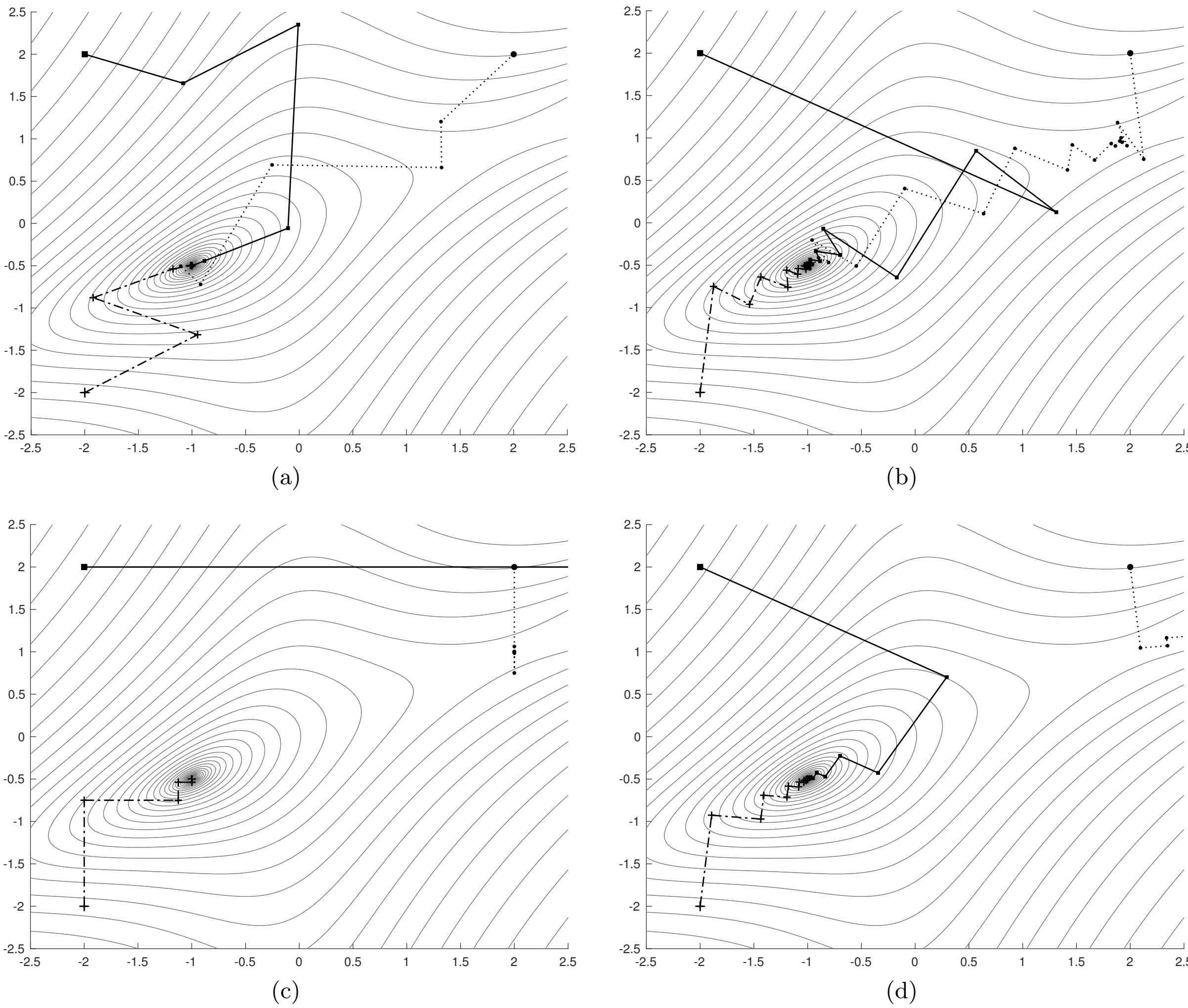


Figure 1: Solution paths of different methods for **Exercise 4**.

Solution:

a) Due to $x \neq x^*, y^*$, we have

$$\nabla f(x)^\top d_\infty = -\left|\frac{\partial f}{\partial x_j}(x)\right|^2 = -\|\nabla f(x)\|_\infty^2 < 0.$$

Hence, d_∞ is a descent direction of f at x.

4 pts in total. 2 pts for correct calculation. 1 pt for $-\|\nabla f(x)\|_\infty^2 < 0$ $(x \neq x^*, y^*)$. 1 pt for final statement / descent direction.

b) We have [(b),1] and [(d),2]. The methods in (b) and (d) use the same initial direction d^0 but with different step sizes. This leaves the gradient methods 1 & 2 as only choice. Since the solution paths in subfigure (d) show perpendicular behavior, we conclude that the method in (d) uses exact line search. The remaining matchings are [(a),3] and [(c),4]. The directions d_∞ can only be along e_1 and e_2. The method in (c) is the only one that matches such behavior. This leaves (a) for Newton's method. (The method in (a) also seems to converge fast within few iterations which is a sign for the fast local convergence of Newton's method).

6 pts in total. 1.5 pts for each correct pairing. No explanation is required.

Exercise 5 (Integer Programming Modeling): (*18* points)

We are considering building some combination of plants of type 1 and 2 in regions a, b, and c. Let x_{ij} be a binary decision variable that takes value 1 if we build a plant of type i in region j and takes value 0 otherwise, where $i \in \{1,2\}$ and $j \in \{a,b,c\}$. Consider each of the following situations independently. Formulate one or more *linear* integer constraints (and objective in part g)) that models the stated condition. Introduce and clearly define additional decision variables if necessary.

a) At least one plant must be built in each region.

b) At least one plant of type 1 must be built.

c) If region a contains a plant of type 1 then region c must also contain a plant of type 1.

d) Region a cannot contain a plant of type 1 and of type 2.

e) Region b can contain a plant of type 2 only if: a plant of type 1 is built in region a and a plant of type 2 is built in region c.

f) Region c contains a plant of type 2 if and only if region a and region b both contain a plant of type 2. You should reformulate it using one or more *linear* constraints.

g) The objective is to maximize the number of regions that contain plants of both types.

Solution:

a) $\sum_{i\in\{1,2\}} x_{ij} \geq 1$ for all $j \in \{a,b,c\}$. (2 pts)

b) $\sum_{j\in\{a,b,c\}} x_{1j} \geq 1$. (2 pts)

c) $x_{1a} \leq x_{1c}$. (2 pts)

d) $x_{1a} + x_{2a} \leq 1$. (3 pts)

e) $x_{2b} \leq x_{1a}, x_{2b} \leq x_{2c}$. (3 pts)

f) $x_{2c} \geq x_{2a} + x_{2b} - 1, x_{2c} \leq x_{2a}, x_{2c} \leq x_{2b}$. (3 pts)

g) max $\sum_{j\in\{a,b,c\}} y_j$, where $y_j \leq x_{1j}, y_j \leq x_{2j}, \forall j \in \{a,b,c\}$. Variables y_j indicates whether there are plants of both types in region j. (It is also okay if students add $y_j \geq x_{1j} + x_{2j} - 1$.) (3 pts)

For part a)–e), any mistake will lead to 0 point. For part f) and g), each expression is worth 1 point.

Exercise 6 (Branch-and-Bound Algorithm): (*16* points)

Figure 2 below depicts an incomplete branch-and-bound tree to maximize a mixed integer program, with numbers next to nodes indicating the LP relaxation objective values. Node 4 has just produced the first incumbent solution. Node a, b, and c are not explored yet.

a) Show which unexplored nodes could be immediately pruned (fathomed) if the integer solution at node 4 corresponds to an objective function value 205. What if the integer solution at node 4 corresponds to an objective function value 210?

b) Determine the best upper bound of the mixed integer program after processing of node 4. You may need to argue by cases about the possible objective values the integer solution at node 4 provides.

c) Assuming the integer solution at node 4 yields an objective value 195, compute the objective value error, i.e., the percentage of the gap relative to the upper bound, in accepting the current best feasible solution as an approximate optimum.

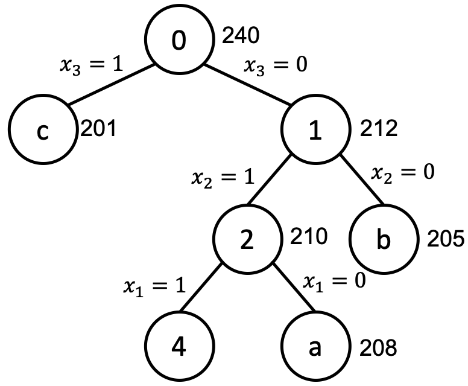


Figure 2: Branch-and-Bound Tree for **Exercise 6**.

Solution:

a) If the integer solution at node 4 has an objective function value 205, nodes b and c can be fathomed, because the LP relaxation objective value of nodes b and c are smaller than 205. If the integer solution at node 4 has an objective function value 205, nodes a, b, and c can all be fathomed for the same suboptimality reason.

5 pts in total. 2 pts for the first statement and 3 pts for the second statement. If they do not provide the reasoning about LP relaxation's suboptimality, take 2 pts off.

b) If the objective value obtained at node 4, z_4^{LP}, is smaller than 208, then the best upper bound is 208, because it is still possible to explore node a and generate an integer solution with the objective value as high as 208. If $z_4^{LP} \in [208, 210]$, then the best upper bound is z_4^{LP} because node a, b, and c are all fathomed.

5 pts in total. 2 pts for the first statement and 3 pts for the second statement.

c) The objective value error can be calculated as follows:

$$\frac{UB - LB}{UB} \times 100\% = \frac{208 - 195}{208} \times 100\% = 6.25\%$$

6 pts in total. 2 pts for correctly identifying UB as 208. 2 pts for correctly identifying LB as 195. 2 pts for correctly obtaining the gap.