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

Prof. Zizhuo Wang · Prof. Haoxiang Yang · Spring Semester 2024

---

# MAT 3007 – Optimization

Final Exam

*Please make sure to present your solutions and answers in a comprehensible way and give explanations of your steps and results!*

- *The exam time is 120 minutes.*
  *Examination Time: 13:30 to 15:30. (No early leave is allowed).*
- *There are six exercises. The total number of achievable points is* 100 *points.*
- *You are allowed to bring two sheets of A4 paper (with arbitrary notes on both sides of it) for your personal use in this exam. You are not allowed to use electronic devices, including a calculator.*
- *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.*
- *Please abide by the honor codes of CUHK-SZ. Any violation of the exam policies will be considered as cheating and reported. Consequences of such a violation include zero points for the midterm exam and corresponding disciplinary actions.*

Good Luck!

**Exercise 1 (Convexity):** (*14* points)

Consider the following function of $\lambda$ and $\mu$ where $a$ and $b$ are two constants.

$$f(\lambda, \mu) = a\lambda \log \lambda - \lambda \log (b + \mu),$$

where the function is defined on $\lambda > 0$ and $\mu > -b$.

(a) (4 pts) For what values of $a$ and $b$ is this function a convex function in $\lambda$? Please give a necessary and sufficient condition.

(We call a function $f(x, y)$ a convex function in $x$ if for any given constant $y_0$, the function $g(x) = f(x, y_0)$ is a convex function.)

(b) (4 pts) For what values of $a$ and $b$ is this function a convex function in $\mu$? Please give a necessary and sufficient condition.

(c) (6 pts) For what values of $a$ and $b$ is this function a convex function in $\lambda$ and $\mu$? Please give a necessary and sufficient condition.

**Answer.**

(a) We first take the second-order derivative of $f$ with respect to $\lambda$. We have $f''(\lambda) = a/\lambda \geq 0$ (2pts) for all $\lambda > 0$ if and only if $a \geq 0$. Therefore, this function is convex in $\lambda$ for all values of $a \geq 0$ and any $b$ (2pts).

(b) We take the second-order derivative of $f$ with respect to $\mu$. We have $f''(\mu) = \frac{\lambda}{(b+\mu)^2} > 0$ (2pts) for all $\lambda > 0$ and $\mu > -b$, regardless of the values of $a$ and $b$. Therefore, this function is convex in $\mu$ for all values of $a$ and $b$ (2pts).

(c) We consider the Hessian matrix of $f$. We have (3pts)

$$H = \begin{pmatrix} \frac{a}{\lambda} & -\frac{1}{b+\mu} \\ -\frac{1}{b+\mu} & \frac{\lambda}{(b+\mu)^2} \end{pmatrix}.$$

This matrix is positive semi-definite if and only if $a \geq 1$. Therefore, this function is convex in $\lambda$ and $\mu$ if and only if $a \geq 1$ (3pts).

**Exercise 2 (KKT Conditions):** (*20* points)

Consider the following optimization problem where $\mathbf{p} = (p_1, ..., p_n)$ is the decision variable. The decision variable corresponds to the probability of $n$ scenarios of an event, which is uncertain to the decision maker. Historically, we have observed $N_1, ..., N_n$ occurrences of scenario $1, ..., n$ respectively, and the likelihood of the historical data is $\prod_{i=1}^n p_i^{N_i}$. Scenario $i$ has reward $r_i$ for the decision maker, and the decision maker wants to calculate the worst-case expected value of the reward when the probability parameters will make the historical observations achieve a likelihood $\gamma > 0$. Therefore, the optimization problem is

$$\begin{aligned} \min \quad & \sum_{i=1}^n r_i p_i \\ \text{s.t.} \quad & \prod_{i=1}^n p_i{}^{N_i} \geq \gamma && (1) \\ & \sum_{i=1}^n p_i = 1 && (2) \\ & p_i \geq 0, \forall i = 1, ..., n. \end{aligned}$$

Please answer the following questions:

(a) (4 pts) Transform the first constraint into a form of $g(\mathbf{p}) \geq 0$ where $g$ is a concave function.

(b) (6 pts) After performing the transformation in part (a), write down the KKT condition of the problem (using the transformed constraint). Please use $\lambda$ to denote the Lagrangian multiplier for the transformation of constraint (1) and $\mu$ to denote the Lagrangian multiplier for constraint (2).

(c) (4 pts) Show that the optimal solution must follow $p_i = \frac{-N_i \cdot \lambda}{r_i + \mu}$ for all $i = 1, \ldots, n$.

(d) (6 pts) What is the maximal value of $\gamma$ so that the optimization problem is feasible?

**Answer.**

(a) We take a logarithm on both side of the first constraint and it will become

$$\sum_{i=1}^n N_i \log p_i \geq \log \gamma \quad \text{(4pts)}$$

Note that $\log p_i$ is a concave function of $p_i$ and thus we can write it as $g(\mathbf{p}) \geq 0$ where $g(\mathbf{p}) = \sum_{i=1}^n N_i \log p_i - \log \gamma$.

(b) To write down the KKT condition, we first construct the Lagrangian:

$$L(\mathbf{p}, \lambda, \mu) = \sum_{i=1}^n r_i p_i + \lambda \left( \sum_{i=1}^n N_i \log p_i - \log \gamma \right) + \mu \left( \sum_{i=1}^n p_i - 1 \right). \quad \text{(2pts)}$$

Now we take the derivative with respect to $p_i$ and obtain the main condition:

$$\frac{\partial L(\mathbf{p}, \lambda, \mu)}{\partial p_i} = r_i + \frac{\lambda N_i}{p_i} + \mu \geq 0.$$

To summarize, we have the following KKT condition:

- Main condition:

$$\frac{\partial L(\mathbf{p}, \lambda, \mu)}{\partial p_i} = r_i + \frac{\lambda N_i}{p_i} + \mu \geq 0, \quad \forall i = 1, ..., n$$

- Dual feasibility condition: $\lambda \leq 0$
- Primal feasibility condition: $\sum_{i=1}^n N_i \log p_i \geq \log \gamma$, $\sum_{i=1}^n p_i = 1$, $p_i \geq 0$ for all $i$.
- Complementarity condition:

$$p_i \cdot \left( r_i + \frac{\lambda N_i}{p_i} + \mu \right) = 0, \quad \forall i = 1, ..., n; \quad \lambda \cdot (\sum_{i=1}^n N_i \log p_i - \log \gamma) = 0$$

Note that two points will be deducted for each wrong condition (except for the primal feasibility condition)

(c) By the complementarity condition, we have $p_i r_i + \lambda N_i + p_i \mu = 0$ is a necessary condition for optimal solution. Thus we must have $p_i = \frac{-N_i \cdot \lambda}{r_i + \mu}$.

(d) To solve for the maximal $\gamma$ for the problem to have a feasible solution, we consider the following optimization problem (3pts):

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^n N_i \log p_i \\ \text{subject to} & \sum_{i=1}^n p_i = 1 \\ & p_i \geq 0, \quad \forall i. \end{array}$$

To solve this problem, we construct the KKT condition, we have the Lagrangian function is

$$\sum_{i=1}^n N_i \log p_i + \lambda \left( \sum_{i=1}^n p_i - 1 \right).$$

By the main condition, we must have $N_i / p_i + \lambda \geq 0$ for all $i$ and by the complementarity condition, we have $N_i + \lambda p_i = 0$. Thus we must have $p_i = -N_i / \lambda$. Meanwhile, by the primal feasibility condition, we have $\sum_{i=1}^n p_i = 1$. Therefore, we have $p_i = \frac{N_i}{\sum_{i=1}^n N_i}$. Thus the maximal $\gamma$ is

$$\gamma^* = \prod_{i=1}^n \left( \frac{N_i}{\sum_{i=1}^n N_i} \right)^{N_i}. \quad \text{(3pts)}$$

If $\gamma > \gamma^*$, then the original problem doesn't have a feasible solution; if $\gamma \leq \gamma^*$, then the original problem has a feasible solution with $p_i = \frac{N_i}{\sum_{i=1}^n N_i}$.

**Note:** The initial formulation worths 3 points (it is also OK to consider to maximize ${p_i}^{N_i}$ in the initial formulation) , and the final solution worth 3 points. All the other intermediate steps won't be credited.

**Exercise 3 (Nonlinear Programming Algorithms):** (*18* points)
We would like to find the minimum of the following function:

$$f(x) = x_1^2 + 2x_2^2 - x_1x_2 - x_1 - 3x_2.$$

We start at the point $x^0 = (0, 0)$.

(a) (4 pts) Derive the gradient descent direction $d^G$. Calculate the step length $\alpha$ if we use the exact line search method.

(b) (4 pts) Derive the descent direction based on Newton's method, i.e., a Newton step, $d^N$.

(c) (3 pts) Suppose we want to minimize the function $f(x)$ within the set of $\{x \mid x_1+2x_2 \leq 1\}$. What is the largest step length $\alpha'$ if we take the gradient descent step $d^G$ from $x^0$? What is the new point after perform such a step, i.e., what is my new $x^1 = x^0 + \alpha' d^G$?

(d) (7 pts) At the obtained point $x^1$ from part (c), suppose we want to perform a project gradient descent iteration, where we project the point $x^1 - \nabla f(x^1)$ to the feasible region $\{x \mid x_1 + 2x_2 \leq 1\}$. Write down the formulation for the projection problem.

**Answer.**

(a) The gradient

$$\nabla f(x) = \begin{bmatrix} 2x_1 - x_2 - 1 \\ 4x_2 - x_1 - 3 \end{bmatrix}$$

and $d^G = -\nabla f(x^0) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ (2pts). The exact line search aims to find the minimizer for

$$\phi(\alpha) = f(x + \alpha d^G) = f(\begin{bmatrix} \alpha \\ 3\alpha \end{bmatrix}) = 16\alpha^2 - 10\alpha.$$

(1pt) This is a convex quadratic function in $\alpha$, and we can use the first order necessary condition to obtain the optimal solution:

$$\phi'(\alpha^*) = 32\alpha^* - 10 = 0,$$

which yields $\alpha^* = \frac{5}{16}$ (1pt).

(b) The Hessian matrix of $f$ is $\nabla^2 f(x) = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$ and its inverse is $(\nabla^2 f(x))^{-1} = \begin{bmatrix} \frac{4}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix}$ (2pts). The Newton's step is (2pts):

$$d^N = -(\nabla^2 f(x^0))^{-1} \nabla f(x^0) = -\begin{bmatrix} \frac{4}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix} \begin{bmatrix} -1 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(If the student gets the numerical value of $d^N$ correct but not providing the Hessian's inverse, give full credits.)

(c) $x^1 = \begin{bmatrix} \alpha \\ 3\alpha \end{bmatrix}$ and the largest $\alpha$ will lead $x^1$ to the boundary $x_1^1 + 2x_2^1 = 1$ (1pt). Therefore, we can solve $\alpha' + 2 \cdot 3\alpha' = 1$ (1pt), $\alpha' = \frac{1}{7}$ and $x^1 = \begin{bmatrix} 1/7 \\ 3/7 \end{bmatrix}$ (1pt)

(d) $-\nabla f(x^1) = -\begin{bmatrix} 2/7 - 3/7 - 1 \\ 12/7 - 1/7 - 3 \end{bmatrix} = \begin{bmatrix} 8/7 \\ 10/7 \end{bmatrix}$ (2pts) and we can obtain $x^1 - \nabla f(x^1) = \begin{bmatrix} 9/7 \\ 13/7 \end{bmatrix}$ (2pts). Therefore, the optimization problem to obtain the projection is (3pts):

$$\begin{aligned} \min_{y \in \mathbb{R}^2} \quad & \frac{1}{2}\left[ (\frac{9}{7} - y_1)^2 + (\frac{13}{7} - y_2)^2 \right] \\ \text{s.t.} \quad & y_1 + 2y_2 \leq 1. \end{aligned}$$

(The objective function can be written as a 2-norm form as well. It is okay to formulate without 1/2.

Another right answer is:

$$\begin{aligned} \min_{d \in \mathbb{R}^2} \quad & -\frac{8}{7}d_1 - \frac{10}{7}d_2 \\ \text{s.t.} \quad & d_1 + 2d_2 \leq 0 \quad (d_1 + 2d_2 = 0 \text{ also okay.}) \\ & \|d\|_2 \leq 1.) \end{aligned}$$

**Exercise 4 (Integer Programming Formulation):** (*16* points)
A cell phone company assembles a new model of cell phones in its factory. Over the next 4 quarters, the company must ship $d_1$, $d_2$, $d_3$ and $d_4$ units of phones to its customers, respectively (the shipments will be made in the end of the quarter). But no more than $M$ units can be assembled in any quarter. There is a fixed cost of $f$ each time the assembly line is setup for production (for the quarter), plus a cost of $c$ per unit assembled. In the beginning, there is no inventory of the phone. The company can only hold up to $L$ units of inventory across quarter (that is, after the shipment of each quarter, the company can have at most $L$ phones left, which could be used for shipment in later quarters). Write an integer *linear* programming formulation for the optimal production plan for the 4 quarters, to minimize the total production cost.

**Answer.** Let $x_1, x_2, x_3, x_4$ denote the production quantity for each quarter. Let $y_1, y_2, y_3, y_4$ be indicator variables of whether to produce in each quarter. That is, $y_i = 1$ means the company produces in quarter $i$, and $y_i = 0$ means otherwise (3pts). The objective will be to minimize (2pts):

$$c \cdot (x_1 + x_2 + x_3 + x_4) + f \cdot (y_1 + y_2 + y_3 + y_4)$$

For the constraint, first, for each quarter, the company can only produce if it pays the setup cost, and the production quantity cannot exceed $M$. Therefore we have (4pts)

$$x_1 \le My_1, \quad x_2 \le My_2, \quad x_3 \le My_3, \quad x_4 \le My_4.$$

Next, the production quantity must satisfy the demand (4pts):

$$x_1 \ge d_1, \quad x_1 + x_2 \ge d_1 + d_2, \quad x_1 + x_2 + x_3 \ge d_1 + d_2 + d_3, \quad x_1 + x_2 + x_3 + x_4 \ge d_1 + d_2 + d_3 + d_4.$$

Also, the remaining inventory at the end of each quarter cannot exceed $L$. That means (3pts, the last constraint is not necessary)

$$x_1 \le d_1 + L, \quad x_1 + x_2 \le d_1 + d_2 + L, \quad x_1 + x_2 + x_3 \le d_1 + d_2 + d_3 + L,$$
$$x_1 + x_2 + x_3 + x_4 \le d_1 + d_2 + d_3 + d_4 + L.$$

Therefore, the entire optimization problem is:

$$\begin{array}{ll} \text{minimize} & c \cdot (x_1 + x_2 + x_3 + x_4) + f \cdot (y_1 + y_2 + y_3 + y_4) \\ \text{subject to} & d_1 \le x_1 \le d_1 + L \\ & d_1 + d_2 \le x_1 + x_2 \le d_1 + d_2 + L \\ & d_1 + d_2 + d_3 \le x_1 + x_2 + x_3 \le d_1 + d_2 + d_3 + L \\ & d_1 + d_2 + d_3 + d_4 \le x_1 + x_2 + x_3 + x_4 \le d_1 + d_2 + d_3 + d_4 + L \\ & 0 \le x_1 \le My_1 \\ & 0 \le x_2 \le My_2 \\ & 0 \le x_3 \le My_3 \\ & 0 \le x_4 \le My_4 \\ & x_1, x_2, x_3, x_4 \in \mathbb{Z}, y_1, y_2, y_3, y_4 \in \{0, 1\}. \end{array}$$

**Note.** If one uses a correct but nonlinear formulation, for example $x_i y_i$ in the formulation, then 5 points will be deducted.

**Exercise 5 (Branch-and-Bound Algorithm):** (*18* points)
Figure 1 contains an incomplete branch-and-bound tree for an integer program, whose objective function is being maximized. The branch-and-bound algorithm selects subproblems (nodes) upon which to branch, using the best-first search rule.

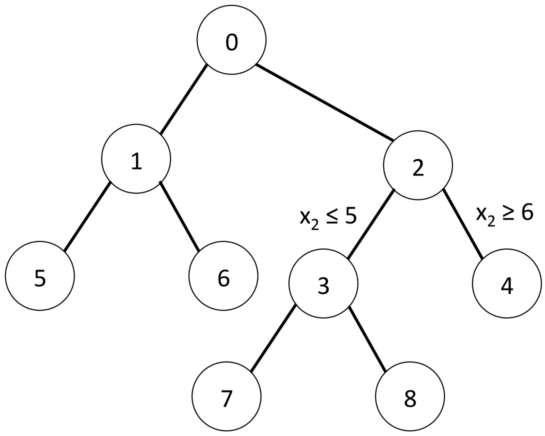


Figure 1: The figure depicts a partial branch-and-bound tree for an integer program.

(a) (14 pts) Let $z_0^{LP}, z_1^{LP}, \ldots, z_8^{LP}$ denote the optimal objective function value of the linear programming relaxation associated with each of the nodes, and assume that the subproblem associated with each node has a feasible linear programming relaxation. *The numbers on the nodes denote the order in which the subproblems were solved in the branch-and-bound algorithm, using the best-first search rule.* Can you infer the following inequalities? Please explain why the inequality holds or does not hold, or the given information is not sufficient to make such inference.

i) (3 pts) $z_1^{LP} \geq z_2^{LP}$

ii) (3 pts) $z_2^{LP} \leq z_3^{LP}$

iii) (4 pts) $z_1^{LP} \geq z_3^{LP}$

iv) (4 pts) $z_5^{LP} \geq z_7^{LP}$

(b) (4 pts) Figure 1 indicates the branch from subproblem #2 involves decision variable $x_2$. What can you infer about the numerical value of $x_2^{LP}$ in the solution to the linear programming relaxation of subproblem #2? Be as specific as possible.

**Answer.**

(a) Conclusions for part i)-(iii) are worth 1pt each. For part iv) the conclusion is worth 2pts.

i) $z_1^{LP} \geq z_2^{LP}$ does not hold because we branch node 2 before node 1 since node 2's children nodes (nodes 3 and 4) are solved before node 1's children nodes (nodes 5 and 6) (1pt). According to the best-first rule, we should branch a node with larger LP relaxation objective value earlier for this maximization problem (1pt).

ii) $z_2^{LP} \leq z_3^{LP}$ does not hold because this is a maximization problem (1pt), the children node contains more constraints than the father node, leading to a smaller objective value (1pt).

iii) $z_1^{LP} \geq z_3^{LP}$ holds because after branching node 2, the remain nodes to be branched are $\{1, 3, 4\}$ (1pt). We branch node 1 before node 3 since node 1's children nodes (nodes 5 and 6) are solved before node 3's children nodes (nodes 7 and 8) (1pt). According to the best-first rule, we should branch a node with larger LP relaxation objective value earlier for this maximization problem (1pt).

iv) $z_5^{LP} \geq z_7^{LP}$ cannot be determined by current information because we do not have the branching information for nodes 5 and 7 (2pts).

(b) $5 < x_2^{LP} < 6$ (1pt for the strict inequalities and 1pt for the numerical correctness). Explanation (2pts): since we are branching $x_2$, $x_2$ should be a fractional number between $\lfloor x_2 \rfloor = 5$ and $\lceil x_2 \rceil = 6$.

**Exercise 6 (Linear and Integer Program):** (*14* points)

Figure 2 shows a network in which each arc has label $c_{ij}$, where $c_{ij}$ is the arc length.

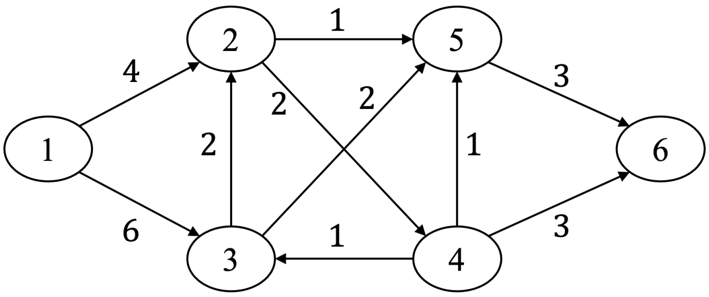


Figure 2: Network with arc labels $c_{ij}$.

(a) (8 pts) Formulate an integer linear program to find the shortest path from node 1 to 6 for the network depicted in Figure 2. Be sure to define your decision variables clearly.

(b) (4 pts) Suppose the optimal objective value to the problem in part (a) is $z^{IP}$. If we relax the constraints that the decision variables are integer-valued, will the objective value to the linear program $z^{LP}$ differ from $z^{IP}$? If so, will $z^{IP} \geq z^{LP}$ or $z^{IP} \leq z^{LP}$?

(c) (2 pts) Suppose you solve the linear program in part (b) and find that the shortest path from node 1 to 6 is $1 \to 2 \to 5 \to 6$. Is the corresponding optimal solution to the model you build in part (b) a degenerate solution? Explain.

**Answer.**

(a) Let the flow on the arc $(i, j)$ be the decision variable $x_{ij}$ (2pts). The IP formulation for the max flow problem is (objective function 2pts; each flow in = flow out constraint 0.5 pt, 3pts in total; integrality constraints 1pt):

$$
\begin{aligned}
\min \quad & 4x_{12} + 6x_{13} + 2x_{24} + x_{25} + 2x_{32} + 2x_{35} + x_{43} + x_{45} + 3x_{46} + 3x_{56} \\
\text{s.t.} \quad & x_{12} + x_{13} && = 1 \\
& x_{24} + x_{25} - x_{12} - x_{32} && = 0 \\
& x_{32} + x_{35} - x_{13} - x_{43} && = 0 \\
& x_{43} + x_{45} + x_{46} - x_{24} && = 0 \\
& x_{56} - x_{25} - x_{35} - x_{45} && = 0 \\
& - x_{46} - x_{56} && = -1 \\
& x_{12}, \ldots, x_{56} \in \{0, 1\}.
\end{aligned}
$$

(b) It will not change because the technology matrix for the max flow problem is totally unimodular (2pt). So all extreme points of the linear programming model in part (b) are integer-valued (1pt). By the LP fundamental theorem, there exists an optimal solution among the extreme points, and thus $z^{LP} = z^{IP}$. (1pt)

(c) It is a degenerate solution (1pt) because there should be six basic elements (columns) but there are only three non-zero values in the decision variable. (1pt)