MAT 3007 Optimization
Final Review

Yuang Chen

School of Data Scienc
The Chinese University of Hong Kong, Shenzhen

July 21, 2025

Final Exam Logistics

- Time and Date: 1:30 - 4:30 pm, July 24 (Thursday)
- Shaw F301 (different from your midterm, seat assignment will be posted soon.)
- Closed-book, closed-notes, no internet, no calculators, no electronics
- Two cheat sheets (double-sided) are allowed.
- 7 problems, 100 points total

Final Exam

- Problem 1 IP Modeling (15 pts)
- Problem 2 Branch-and-Bound Algorithm (10 pts)
- Problem 3 Optimality Conditions (15 pts)
- Problem 4 Convexity (15 pts)
- Problem 5 KKT Conditions (15 pts)
- Problem 6 Algorithms (15 pts)
- Problem 7 Short Answers (15 pts)

Problem 1 Modeling (15 points)

If $f(x_1, x_2, ..., x_n) > a$, then $g(x_1, x_2, ..., x_n) \geq b$.

$$\begin{cases} f(x_1, x_2, ..., x_n) \leq a + My \\ g(x_1, x_2, ..., x_n) \geq b - M(1-y) \\ y \in \{0, 1\} \end{cases}$$

Either $f(x_1, x_2, ..., x_n) \leq a$ or $g(x_1, x_2, ..., x_n) \leq b$

disjunctive
at least ONE

$$\begin{cases} f(x_1, x_2, ..., x_n) \leq a + My \\ g(x_1, x_2, ..., x_n) \leq b + M(1-y) \\ y \in \{0, 1\} \end{cases}$$

Problem 1 Modeling (15 pts)

(a) x_i for $i = 1, 2, 3$ are continuous nonnegative variables. Write a set of constraints to model the requirement that:

$$|2x_1 - x_2 - x_3| \geq 2$$

by introducing an additional binary variable.

(b) x_1 and x_2 are integer variables. Write a set of constraints to model the requirement that: either $x_1 + x_2 \leq 10$ or $2x_1 - x_2 \geq 5$ but not both by introducing an additional binary variable.

(c) If at least 3 of 4 events are selected from events A, B, C, D, then no more than 2 of 3 events can be selected from events E, F, G.

1. (a) $|2x_1 - x_2 - x_3| \geq 2$

$\Downarrow$

$2x_1 - x_2 - x_3 \geq 2$ OR $2x_1 - x_2 - x_3 \leq -2$ Only one satisfied

$y=0$ $y=1$

$$\begin{cases} 2x_1 - x_2 - x_3 \geq 2 - My \\ 2x_1 - x_2 - x_3 \leq -2 + M(1-y) \\ y \in \{0, 1\} \end{cases}$$

(b) $y=1$ $y=0$

$\begin{cases} x_1 + x_2 \leq 10 \\ 2x_1 + x_2 < 5 \end{cases}$ or $\begin{cases} x_1 + x_2 > 10 \\ 2x_1 + x_2 \geq 5 \end{cases}$

Since $x_1, x_2 \in \mathbb{Z}$

$\begin{cases} x_1 + x_2 \leq 10 \\ 2x_1 + x_2 \leq 4 \end{cases}$ or $\begin{cases} x_1 + x_2 \geq 11 \\ 2x_1 + x_2 \geq 5 \end{cases}$

(4, 9)

$y=1$ $y=0$

$$\begin{cases} x_1 + x_2 \leq 10 + M(1-y) \\ 2x_1 + x_2 \leq 4 + M(1-y) \\ x_1 + x_2 \geq 11 - My \\ 2x_1 + x_2 \geq 5 - My \\ y \in \{0, 1\} \end{cases}$$

$$X_i = \begin{cases} 1 & \text{if event } i \text{ is selected} \\ 0 & \text{o.w.} \end{cases}$$

$>$

(c) If $X_A + X_B + X_C + X_D \geq \cancel{3}$

Then $X_E + X_F + X_G \leq 2$

X_i's are binary variables

If $X_A + X_B + X_C + X_D > 2.8$

Then $X_E + X_F + X_G \leq 2$

$$\begin{cases} X_A + X_B + X_C + X_D \leq 2.8 + My \\ X_E + X_F + X_G \leq 2 + M(1-y) \\ y \in \{0, 1\} \end{cases}$$

Problem 2 Branch-and-Bound Algorithm (10 pts)

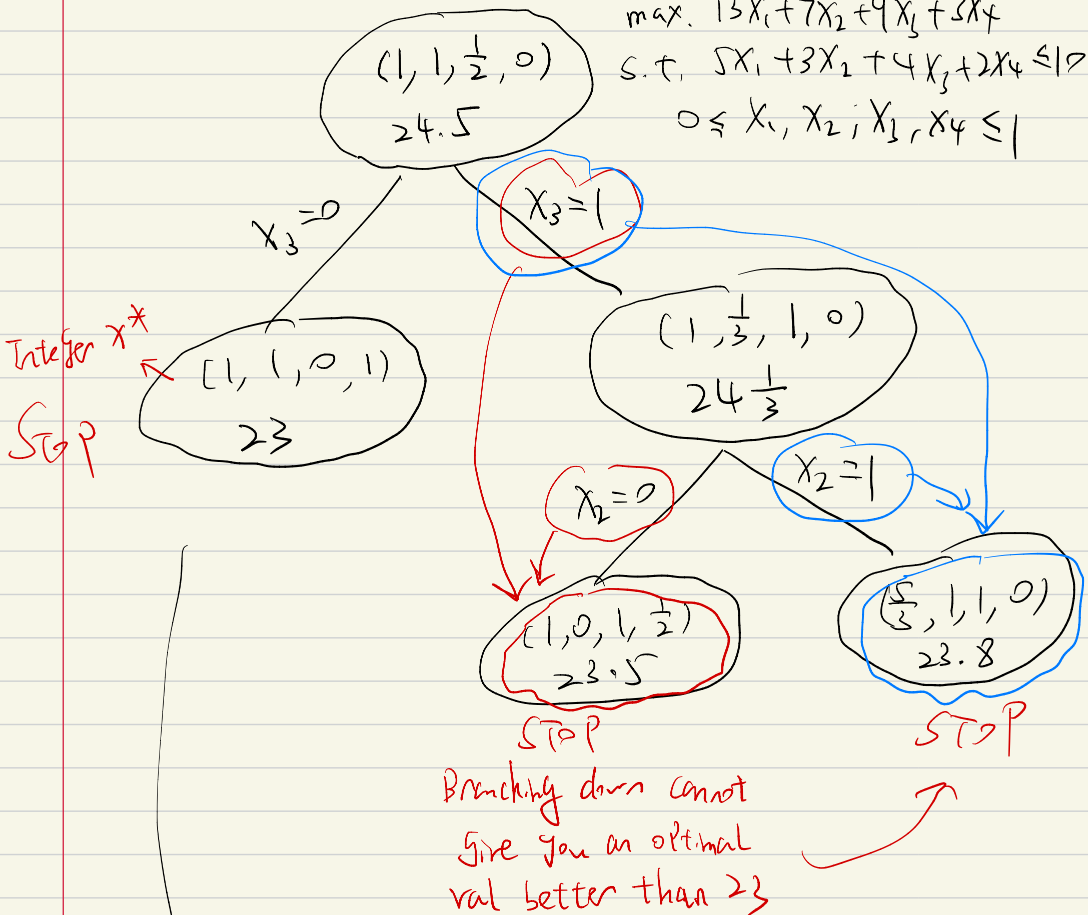
Consider the following knapsack problem:

$$\begin{aligned} \text{maximize} \quad & 13x_1 + 7x_2 + 9x_3 + 3x_4 \\ \text{subject to} \quad & 5x_1 + 3x_2 + 4x_3 + 2x_4 \leq 10 \\ & x_1, x_2, x_3, x_4 \in \{0, 1\}. \end{aligned}$$

Use the branch-and-bound method to solve it (draw the branch-and-bound tree and mark the results on each node).

Profit/weight ration $\frac{13}{5} > \frac{7}{3} > \frac{9}{4} > \frac{3}{2}$

$$\max.\ 13x_1 + 7x_2 + 9x_3 + 3x_4$$
$$\text{s.t. } 5x_1 + 3x_2 + 4x_3 + 2x_4 \le 10$$
$$0 \le x_1, x_2, x_3, x_4 \le 1$$



true optimal $x^* = (1, 1, 0, 1)$

$v^* = 23$

Problem 3 Optimality Conditions (15 pts)

Consider the function $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$,

$$f_\alpha(x) := \alpha x_1^2 + x_2^2 - 2x_1x_2 - 2x_2,$$

where $\alpha \in \mathbb{R}$ is a scalar.

(a) Find the stationary points (in case they exist) of f_α for each value of α.

(b) For each stationary point x^* in part (a), determine whether x^* is a local maximizer or a local minimizer or a saddle point of f_α.

(c) For which values of α can f_α have a global minimizer?

(a) $$\nabla f_a(X) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2a x_1 - 2x_2 \\ 2x_2 - 2x_1 - 2 \end{bmatrix}$$

$$\nabla f_a(X) = 0$$

$$\begin{cases} 2a x_1 = 2x_2 \\ 2x_2 = 2x_1 + 2 \end{cases} \Rightarrow \begin{cases} x_1^* = \frac{1}{a-1} \\ x_2^* = \frac{a}{a-1} \end{cases} \quad \underset{\sim\sim}{a \neq 1}$$

(b) $$\nabla^2 f_a(X) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2a & -2 \\ -2 & 2 \end{bmatrix}$$

$$\det(\nabla^2 f_a(x)) = \begin{vmatrix} 2a & -2 \\ -2 & 2 \end{vmatrix} = 4a - 4 = 4(a-1)$$

- If $a < 1$, $\det(\nabla^2 f_a(x^*)) < 0$

$$\det(\nabla^2 f_a(x^*)) = \lambda_1 \lambda_2 < 0$$

One positive and one negative eigenvalue.

So $\nabla^2 f_a(x^*)$ is indefinite.

x^* is a saddle point.

- If $\alpha > 1$, $\det(\nabla^2 f_\alpha(x^*)) > 0$

$\hookrightarrow \lambda_1 \lambda_2 > 0$

$$\operatorname{tr}(\nabla^2 f_\alpha(x^*)) = 2\alpha + 2 > 0$$

$$= \lambda_1 + \lambda_2 > 0$$

Two eigenvalues are both Positive.

So $\nabla^2 f_\alpha(x^*)$ is Positive-definite

x^* is a local min.

(c) f_α can have global min if $\alpha > 1$.

Problem 4 Convexity (15 pts)

- (a) Verify whether the following set is convex or not:

$$X = \{x \in \mathbb{R} : \alpha \leq \sqrt{x} \leq \beta\}, \quad \alpha \in \mathbb{R}, \quad \beta \geq 0, \quad \alpha \leq \beta.$$

- (b) Verify $f(x) := \frac{1}{2}\|Ax - b\|_2^2 + \mu\|Lx\|_\infty$ convex is a convex function or not, where $\|\cdot\|_\infty$ denotes the maximum norm.
- (c) Verify $f(x) := \frac{1}{2}\|Ax - b\|_2^2 + \mu\|Lx\|_\infty + \lambda\|Qx\|_0$ not convex is a convex function or not, where $\|\cdot\|_\infty$ denotes the maximum norm, $\|x\|_0$ counts the number of non-zero elements in x.
- (d) Verify $f(x_1, x_2) = -\ln x_1 - \ln x_2$ is convex or not on the interval $\{(x_1, x_2) : x_1 \cancel{\geq} 0, x_2 \cancel{\geq} 0\}$.

  $>$ $>$

(a) $\forall x, y \in \mathcal{X}$, $\begin{cases} \alpha \le \sqrt{x} \le \beta \Rightarrow (\max\{0, \alpha\})^2 \le x \le \beta^2 \\ \alpha \le \sqrt{y} \le \beta \Rightarrow (\max\{0, \alpha\})^2 \le y \le \beta^2 \end{cases}$

$\forall \lambda \in [0, 1]$

$$\begin{array}{rcccl} & \lambda(\max\{0, \alpha\})^2 \le & \lambda x & \le \lambda \beta^2 \\ + & (1-\lambda)(\max\{0, \alpha\})^2 \le & (1-\lambda) y & \le (1-\lambda)\beta^2 \\ \hline & (\max\{0, \alpha\})^2 \le & \lambda x + (1-\lambda) y & \le \beta^2 \end{array}$$

$$\alpha \le \sqrt{\lambda x + (1-\lambda) y} \le \beta$$

So $\lambda x + (1-\lambda) y \in \mathcal{X}$

$\mathcal{X}$ is a convex set.

(b) Prove $\|x\|_\infty$ is convex.

$\|x\|_\infty = \max\{x_1, \cdots, x_n\}$

$\forall \lambda \in [0, 1]$

$$\|\lambda x + (1-\lambda) y\|_\infty = \max_i \{\lambda x_i + (1-\lambda) y_i\}$$

j takes optimal

$$\lambda x_j + (1-\lambda) y_j$$

$$\le \lambda \max_i \{x_i\} + (1-\lambda) \max_i \{y_i\}$$

$$= \lambda \|x\|_\infty + (1-\lambda) \|y\|_\infty$$

(c) Prove $\|X_0\|$ is not convex

$\|X\|_0 = \#$ of nonzeros in X

$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\lambda = 0.5$

$\| \lambda x + (1-\lambda) y \|_0 = \| \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \| = 2$

$\lambda \|X\|_0 + (1-\lambda) \|y\|_0 = 0.5 \times 1 + 0.5 \times 1 = 1$

(d) $f(x_1, x_2) = -\ln x_1 - \ln x_2$

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

$\det(\nabla^2 f(x)) = \frac{1}{x_1^2 x_2^2} > 0$

$\operatorname{tr}(\nabla^2 f(x)) = \frac{1}{x_1^2} + \frac{1}{x_2^2} > 0$

$\nabla^2 f(x) \succ 0$

So $f(x)$ is convex.

Problem 5 KKT Conditions (15 pts)

Consider the problem

$$
\begin{aligned}
\min_{x\in\mathbb{R}^2} \quad & 4x_1^2 + x_2^2 - x_1 - 2x_2 \quad = f(x) \\
\text{s.t.} \quad & \underbrace{2x_1 + x_2 \le 1}_{g_1(x)}, \quad \underbrace{x_1^2 \le 1}_{g_2(x)}.
\end{aligned}
$$

λ_1 λ_2

(a) Show the above problem is a convex optimization problem.

(b) Show that Slater's condition is satisfied for the above problem.

(c) Derive the KKT conditions for the above problem and find all KKT points.

(d) Does this problem have a unique global solution? Briefly explain your answer!

(a) $\nabla f(x) = \begin{bmatrix} 8x_1 - 1 \\ 2x_2 - 2 \end{bmatrix}$

$\nabla^2 f(x) = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$

$\nabla^2 f(x) \succ 0 \Rightarrow f(x)$ is convex

$g_1(x)$ and $g_2(x)$ are convex.

So it's convex OPT.

(b) $x = (0, 0)$ $\begin{cases} 0 + 0 < 1 \\ 0^2 < 1 \end{cases}$

Slater's condition satisfied

(c) KKT Conditions:

Main. $\nabla f(x) + \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x)$

$= \begin{bmatrix} 8x_1 - 1 + 2\lambda_1 + 2x_1\lambda_2 \\ x_2 - 2 + \lambda_1 \end{bmatrix}$

$\begin{cases} 8x_1 - 1 + 2x_1 + 2x_1\lambda_2 = 0 \\ x_2 - 2 + \lambda_1 = 0 \end{cases}$ r

Complementary slackness

$$\lambda_1 (2x_1 + x_2) = 0$$
$$\lambda_2 \, x_1^2 = 0$$

Primal & dual feasible

$2x_1 + x_2 \le 1 \qquad \lambda_1, \lambda_2 \ge 0$

$x_1^2 \le 1$

Case I: $\lambda_1 = \lambda_2 = 0$

$x_1 = \frac{1}{8}, \; x_2 = 1 \Rightarrow$ infeasible

Case II $\lambda_2 = 0 \quad 2x_1 + x_2 = 0$

$x_1 = \frac{1}{16}, \; x_2 = \frac{7}{8} \quad \lambda_1 = \frac{1}{4}$

$\rightarrow$ KKT Point

Case III: $\lambda_1 = 0, \; x_1^2 = 0$

$x_1 = 0$

Main second $\rightarrow$ $8 \cdot 0 - 1 + 2 \cdot 0 + 2 \cdot 0 = -1 \neq 0$

$\Rightarrow$ not Possible

Case IV $2x_1 + x_2 = 0, \; x_1^2 = 0$

$\Rightarrow$ not Possible

Single KKT Pt: $\left(\frac{1}{16}, \frac{7}{8}\right)$

(d) Since the Problem is a convex OPT and slater's condition is satisfied,

So $x^* = (\frac{1}{16}, \frac{7}{8})$ is global min.

- Convex + slater's (Strong duality)

KKT $\Leftrightarrow$ optimal

Problem 6 Algorithm (15 pts)

Let us consider the following least-squares optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x) := \|Ax - b\|^2, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given.

(a) Suppose that $A^\top A \in \mathbb{R}^{n \times n}$ has full rank and we apply Newton's method for solving problem (1). We start from some $x^0 \in \mathbb{R}^n$. Compute the first iterate x^1 using Newton's method. What property does x^1 have?

(b) Suppose that $m \geq 2$. In this case, the mapping f in (1) can be written as $f(x) = \sum_{i=1}^m f_i(x)$ with $f_i(x) = (a_i^\top x - b_i)^2$, where $a_i^\top \in \mathbb{R}^{1 \times n}$ is the i-th row of A and b_i is the i-th element of b.
Assume that $x^k = 0 \in \mathbb{R}^n$ at the k-th iteration. Is $-\nabla f_i(x^k)$ always a descent direction of f at x^k? If yes, justify your answer. If no, provide a suitable counterexample.

$-\nabla f_i(0)$

(a) $x^1 \leftarrow x^0 - [\nabla^2 f(x^0)]^{-1} \nabla f(x^0)$

$\nabla f(x) = 2A^T(Ax-b)$

$\nabla^2 f(x) = 2A^TA$

$$x^1 = x^0 - (2A^TA)^{-1}(2A^T(Ax^0-b))$$

$$= x^0 - (\cancel{(A^TA)^{-1}A^TA}x^0 + (A^TA)^{-1}A^Tb)$$

$$= x^0 - x^0 + (A^TA)^{-1}A^Tb$$

$$= (A^TA)^{-1}A^Tb$$

$\left\{\begin{array}{l} \nabla f(x^1) = 0 \\ \text{convex opt} \end{array}\right. \Rightarrow x^1$ is global optimal

(b) descent direction: $\{d : (\nabla f(x))^T d < 0\}$

$$-(\nabla \underset{\smile}{f_i}(0))^T \nabla \underset{\smile}{f}(0)$$

scalar vector

$\nabla f_i(x) = 2(a_i^T x - b_i) a_i$

$\nabla f(x) = 2A^T(Ax-b)$

$\nabla f_i(0) = -2b_i a_i$

$\nabla f(0) = -2A^Tb$

$$-(\nabla f_i(0))^T \nabla f(0)$$
$$= -(-2b_i a_i)^T(-2A^T b)$$
$$= -4 b_i a_i^T (A^T b) < 0 \ ?$$

not always

May not

Counter example:

$$A = \begin{bmatrix} -a_1^T- \\ --a_1^T- \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad b = \begin{bmatrix} -b_1 \\ b_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\to = 0 \not< 0$$

So $\nabla f_i(x^k)$ may not be a descent direction at $x^k = 0$

Problem 6 Algorithm (15 pts)

(c) Let A and b be given via

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Suppose that we apply the gradient descent method for solving problem (4) for this specific choice of A and b. We use backtracking line search with parameters $\sigma, \gamma \in (0, 1)$ for choosing the step size α_k. Assume that $x^k = (0, 0)^\top$ at the k-th iteration. What is the range of γ so that $\alpha_k = 1$ will be chosen by the line search procedure?

Remark: Backtracking line search procedure is to determine the step size α_k as the largest element in $\{1, \sigma^2, \sigma^3, \dots\}$ such that

$$f(x^k + \alpha_k d^k) - f(x^k) \le \gamma \alpha_k \nabla f(x^k)^\top d^k.$$

(c) $f(x^k + \alpha_k d^k) - f(x^k)$

$$= \|A(x^k + \alpha_k d^k) - b\|^2 - \|Ax^k - b\|^2$$

$$= \|Ax^k - b + \alpha_k A d^k\|^2 - \|Ax^k - b\|^2$$

$$= \cancel{\|Ax^k - b\|^2} + 2\alpha_k (Ax^k - b)^T d^k + \alpha_k^2 \|Ad^k\|^2 - \cancel{\|Ax^k - b\|^2}$$

$$= 2\alpha_k (Ax^k - b)^T d^k + \alpha_k^2 \|Ad^k\|^2$$

$x^k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\alpha_k = 1$,

$d^k = -\nabla f(0) = -2A^T b = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$

$$= -8\alpha_k + 5\alpha_k^2$$

$$\gamma \alpha_k \nabla f(x^k) d^k = \gamma \alpha_k \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = -8\gamma\alpha_k$$

$-8\alpha_k + 5\alpha_k^2 \le -8\gamma\alpha_k$ when $\alpha_k = 1$

$$\Rightarrow \quad 0 < \gamma \le \frac{3}{8}$$

Problem 7 Short Answer Questions (15 pts)

- Three question (5 pts each). Two of them will be something we discussed in class.
- (a) What are the update equations in each iteration for gradient descent and Newton's method for a maximization problem?
- (b) Is Lagrangian dual problem always a convex optimization problem?
- (c) Given the class's notations, if x^*, λ^*, and μ^* are optimal solutions to the primal and dual problems and satisfy the KKT conditions, does this imply that the primal problem must be a convex optimization problem?

$\max.\ f(x)$

(a) $x^{k+1} \leftarrow x^k + \alpha_k \nabla f(x^k)$

$x^{k+1} \leftarrow x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$

(b) Dual: $\max\limits_{\lambda \geq 0, \mu} \mathcal{L}(\lambda, \mu)$

$= \max\limits_{\lambda \geq 0, \mu} \min\limits_{x}.\ L(x, \lambda, \mu)$

$= \max\limits_{\lambda \geq 0, \mu} \min\limits_{x}.\ f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x)$

Linear in λ and μ

max linear $\Rightarrow$ Convex OPT

(c) Counter example

$\min.\ -x^2$

s.t. $x = 0$ not Convex OPT.

$x^* = 0$, satisfy KKT, global optimal

- Thank you for taking the optimization class with me this semester!
- People often say CUHK-Shenzhen is an "OR university" because we have many outstanding professors conducting cutting-edge research in operations research.
- CUHK-Shenzhen offers many advanced optimization courses through the School of Data Science. Optimization also plays a key role in fields such as machine learning, statistics, computer algorithms, automatic control, signal processing, energy systems, financial engineering, and more.
- Good luck with your future studies at CUHK-Shenzhen! I am the director of the DSBDT program in the School of Data Science. If you have any questions about the major, I'm always happy to help.