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

Andre Milzarek · Junfeng Wu · Junchi Yang
Spring Semester 2025

MAT 3007 – Optimization

Solutions — Final Exam — Sample

Problem 1 (Convexity of Functions): (*18* points)
Investigate whether each of the following functions is convex, concave, or neither convex nor concave. Justify and explain your answer!

a) $f : \mathbb{R}^2 \to \mathbb{R}$ and $f(\boldsymbol{x}) = x_1^2 - x_1x_2 + (x_2 - 100)^2$.

b) $f : \mathbb{R} \to \mathbb{R}$ and $f(x) = e^{|x|}$.

c) $f : \mathbb{R}^n \to \mathbb{R}$ and $f(\boldsymbol{x}) = \sum_{i=1}^m \max\{0, 1 - b_i \cdot \boldsymbol{a}_i^\top \boldsymbol{x}\} + \frac{\lambda}{2}\|\boldsymbol{x}\|^2$, where $\lambda > 0$, $b_i \in \{-1, +1\}$, and $\boldsymbol{a}_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, are given constants and data points.

d) $f : \mathbb{R} \to \mathbb{R}$ and $f(x) = \sqrt{x} + \frac{\alpha}{2}x^2$ where $\alpha > 0$ is a constant.

Solution :

a) Convex (2pts). The function can be rewritten as $f(\boldsymbol{x}) = \frac{1}{2}(x_1 - x_2)^2 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 200x_2 + 100^2$. The first term is convex due to the composition of a convex function with a linear function. The overall function is convex due to nonnegative linear combination of convex (and linear combination of linear) functions (2pts).

b) Convex (2pts). For any $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$, we have

$$e^{|\lambda x + (1-\alpha)y|} \le e^{\lambda|x| + (1-\alpha)|y|} \le \lambda e^{|x|} + (1 - \alpha)e^{|y|},$$

where the first inequality is from convexity of $z \mapsto |z|$ and non-decreasing property of $z \mapsto e^z$ and the second inequality is due to the convexity $z \mapsto e^z$ (2pts).

c) Convex (2pts). $z \mapsto \max\{0, z\}$ is convex, and hence $\max\{0, 1 - b_i \cdot \boldsymbol{a}_i^\top \boldsymbol{x}\}, \forall 1 \le i \le m$ are convex due to the composition of a convex function with a linear function. We conclude that f is convex as it a a nonnegative linear combination of convex functions (2pts).

d) Neither convex nor concave (2pts). Taking second-order derivative gives

$$f''(x) = -\frac{1}{4}\frac{1}{x^{3/2}} + \alpha.$$

When $x \to 0$, we have $f''(x) < 0$, while $f''(x) > 0$ when $x \to +\infty$ (3pts). Argument like "convex + concave produces neither convex nor concave" is not a valid justification.

Problem 2 (Optimality Conditions): (*22* points)
Consider the unconstrained optimization problem

$$\text{minimize}_{\boldsymbol{x}\in\mathbb{R}^2}\ f(\boldsymbol{x}) := (x_1 - x_2^2)^2 + x_1 x_2^2 - 3x_1 \tag{1}$$

and let $\boldsymbol{x}^* = (2,1)^\top$ be a given point.

a) Calculate the gradient and Hessian of the objective function f.

b) Compute all stationary points of the minimization problem (1). For each of the stationary points, determine whether it is a local maximizer, local minimizer, or saddle point and explain your answer.

c) We consider the following constrained variant of problem (1):

$$\text{minimize}_{\boldsymbol{x}\in\mathbb{R}^2}\ f(\boldsymbol{x}) = (x_1 - x_2^2)^2 + x_1 x_2^2 - 3x_1 \quad \text{subject to} \quad x_2 \geq \frac{1}{3}x_1 + \frac{1}{3}. \tag{2}$$

Write down the KKT conditions for (2) and show that $\boldsymbol{x}^*$ is a KKT point of this problem.

d) Prove that $\boldsymbol{x}^*$ is an optimal (global) solution of problem (2).

Solution :

a) The gradient and Hessian of f are given by:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2(x_1 - x_2^2) + x_2^2 - 3 \\ -4(x_1 - x_2^2)x_2 + 2x_1x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2^2 - 3 \\ 2x_2(2x_2^2 - x_1) \end{pmatrix}, \quad \nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 2 & -2x_2 \\ -2x_2 & 12x_2^2 - 2x_1 \end{pmatrix}.$$

3 pts in total. 1.5 pts for correct gradient; 1.5 pts for correct Hessian (-0.5 for errors; depending whether this leads to strong simplifications).

b) We first determine all stationary points of f. We consider two cases.

Case 1: $x_2 = 0$. We then obtain $2x_1 = 3$ and hence, the point $\bar{\boldsymbol{x}} = (1.5, 0)^\top$ is a stationary point.

Case 2: $x_2 \neq 0$ and $x_1 = 2x_2^2$. We use this expression in the first equation to obtain $3x_2^2 - 3 = 0$. This yields $x_2 = \pm 1$ and $x_1 = 2$. Thus, $\bar{\boldsymbol{y}} = \boldsymbol{x}^* = (2,1)^\top$ and $\bar{\boldsymbol{z}} = (2,-1)^\top$ are two more stationary points.

We now check the Hessian $\nabla^2 f$ for the three stationary points:

$$\nabla^2 f(\bar{\boldsymbol{x}}) = \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix}, \quad \nabla^2 f(\bar{\boldsymbol{y}}) = \begin{pmatrix} 2 & -2 \\ -2 & 8 \end{pmatrix}, \quad \nabla^2 f(\bar{\boldsymbol{z}}) = \begin{pmatrix} 2 & 2 \\ 2 & 8 \end{pmatrix}.$$

Clearly, $\nabla^2 f(\bar{\boldsymbol{x}})$ is indefinite and thus, $\bar{\boldsymbol{x}}$ is a saddle point. Furthermore, it holds that $\text{tr}(\nabla^2 f(\bar{\boldsymbol{y}})) = \text{tr}(\nabla^2 f(\bar{\boldsymbol{z}})) = 10$ and $\det(\nabla^2 f(\bar{\boldsymbol{y}})) = \det(\nabla^2 f(\bar{\boldsymbol{z}})) = 16 - 4 = 12$. Consequently, $\nabla^2 f(\bar{\boldsymbol{y}})$ and $\nabla^2 f(\bar{\boldsymbol{z}})$ are strict local minimizer of f.

8 pts in total. 1 pt for each correct stationary point. 1 pt for $\nabla^2 f$. 1 pt for checking and verifying definiteness of $\nabla^2 f$ for each stationary point. 1 pt for $\bar{\boldsymbol{x}}$ = saddle point. 1pt for $\bar{\boldsymbol{y}}$, $\bar{\boldsymbol{z}}$ = local minimizers.

c) We define $g(\boldsymbol{x}) := \frac{1}{3}x_1 - x_2 + \frac{1}{3}$. The KKT conditions are given by

$$\nabla_x L(\boldsymbol{x}, \lambda) = \nabla f(\boldsymbol{x}) + \nabla g(\boldsymbol{x})\lambda = \begin{pmatrix} 2x_1 - x_2^2 - 3 \\ 2x_2(2x_2^2 - x_1) \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ -1 \end{pmatrix} \lambda = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$\lambda \geq 0$, $g(\boldsymbol{x}) \leq 0$, and $\lambda g(\boldsymbol{x}) = 0$. For the point $\boldsymbol{x}^* = (2,1)^\top$, it follows $g(\boldsymbol{x}^*) = 0$ and $\nabla f(\boldsymbol{x}^*) = 0$ (by part b)). Hence, the KKT conditions are satisfied with $\lambda = 0$ and $\boldsymbol{x}^*$ is a KKT point of (2).

4 pts in total. 2 pts for complete and correct KKT conditions (-0.5 for missing parts / errors). 2 pts for verification that $\boldsymbol{x}^*$ is KKT point (feasibility + $\lambda = 0$).

d) The function g is linear and convex. Hence, the feasible set $X := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq 0\}$ defines a convex set. If problem (2) is a convex optimization problem, then the KKT point $\boldsymbol{x}^*$ will be automatically a global solution. Hence, we only need to show that the objective function f is convex on X.

For all $\boldsymbol{x} \in X$, it holds that $x_1 \leq 3x_2 - 1$,

$$\begin{aligned} \det(\nabla^2 f(\boldsymbol{x})) &= 24x_2^2 - 4x_2^2 - 4x_1 = 20x_2^2 - 4x_1 \\ &= (12x_1^2 - 12x_2 + 3) + 8x_2^2 + 1 \geq 3(2x_2 - 1)^2 + 1 \geq 1 \end{aligned}$$

and $\mathrm{tr}(\nabla^2 f(\boldsymbol{x})) = 12x_2^2 - 2x_1 + 2 \geq 12x_2^2 - 6x_2 + 4 = (9x_2 - 1)^2 + 3x_2^2 + 1 \geq 1$. This shows that the Hessian $\nabla^2 f(\boldsymbol{x})$ is positive (semi)definite for all $\boldsymbol{x} \in X$. Hence, f is convex on X and $\boldsymbol{x}^*$ is a global solution of (2).

7 pts. 1 pt for convexity of X; 3 pts for proof strategy (convex problem; KKT points are global solutions). 3 pts for checking convexity of f on X (different possibilities).

Problem 3 (True or False): (*12* points)

State whether each of the following statements is *true* or *false*. For each part, only your answer, which should be one of *true* or *false*, will be graded. Explanations are not required and will not be read.

a) Consider the nonlinear program $\min_{\boldsymbol{x} \in X} f(\boldsymbol{x})$ with linear constraints $X := \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$. Let $\boldsymbol{x}^*$ be a local solution of this problem. Then $\boldsymbol{x}^*$ satisfies the KKT conditions.

b) Let $f, h : \mathbb{R}^n \to \mathbb{R}$ be convex functions and let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ be given. Then, the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \quad h(\boldsymbol{x}) = \boldsymbol{0} \end{array}$$

is a convex optimization problem.

c) We consider the standard integer program

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \boldsymbol{x} \in \mathbb{Z}^n. \end{array} \tag{3}$$

If the integer problem (3) is infeasible, then its corresponding LP relaxation is infeasible as well.

d) Let $\boldsymbol{A}$ be a given 2×2 matrix and suppose that every component of $\boldsymbol{A}$ is either -1, 0, or $+1$. Then, $\boldsymbol{A}$ is totally unimodular.

Solution :

a) **True.** Linear constraints are a constraint qualification. (3 pts)

b) **False.** The function $h(x) = x^2 - 1$ is convex. However, the set $\{x : h(x) = 0\} = \{-1, +1\}$ is clearly nonconvex (3 pts)

c) **False.** Take $A = 1$ and $b = \frac{1}{2}$. Then, we have $x = \frac{1}{2} \notin \mathbb{Z}$, i.e., the IP is infeasible. The corresponding LP relaxation is clearly feasible. (3 pts)

d) **False.** Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Then, we have $\det(\boldsymbol{A}) = -1 - 1 = -2 \notin \{-1, 0, 1\}$. Hence, $\boldsymbol{A}$ is not TU. (3 pts)

Problem 4 (Algorithms for Unconstrained Problems): (*16* points)

Let us consider the following least-squares optimization problem:

$$\text{minimize}_{\boldsymbol{x} \in \mathbb{R}^n} \ f(\boldsymbol{x}) := \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2, \tag{4}$$

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

a) Suppose that $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n \times n}$ has full rank and we apply Newton's method for solving problem (4). We start from some $\boldsymbol{x}^0 \in \mathbb{R}^n$. Compute the first iterate $\boldsymbol{x}^1$ using Newton's method. What property does $\boldsymbol{x}^1$ have?

b) Suppose that $m \geq 2$. In this case, the mapping f in (4) can be written as $f(\boldsymbol{x}) = \sum_{i=1}^m f_i(\boldsymbol{x})$ with $f_i(\boldsymbol{x}) = (\boldsymbol{a}_i^\top \boldsymbol{x} - b_i)^2$, where $\boldsymbol{a}_i^\top \in \mathbb{R}^{1 \times n}$ is the i-th row of $\boldsymbol{A}$ and b_i is the i-th element of $\boldsymbol{b}$.

   Assume that $\boldsymbol{x}^k = 0 \in \mathbb{R}^n$ at the k-th iteration. Is $-\nabla f_i(\boldsymbol{x}^k)$ always a descent direction of f at $\boldsymbol{x}^k$? If yes, justify your answer. If no, provide a suitable counterexample.

   Hint: You can fix $i = 1$ without loss of generality.

c) Let $\boldsymbol{A}$ and $\boldsymbol{b}$ be given via

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

   Suppose that we apply the gradient descent method for solving problem (4) for this specific choice of $\boldsymbol{A}$ and $\boldsymbol{b}$. We use backtracking line search with parameters $\sigma, \gamma \in (0, 1)$ for choosing the step size α_k. Assume that $\boldsymbol{x}^k = (0, 0)^\top$ 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 $\{1, \sigma^2, \sigma^3, \ldots\}$ such that $f(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}^k) - f(\boldsymbol{x}^k) \leq \gamma \alpha_k \nabla f(\boldsymbol{x}^k)^\top \boldsymbol{d}^k$.

Solution :

a) We first compute $\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$ and $\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}^\top\boldsymbol{A}$. By Newton's iteration, we have $\boldsymbol{x}^1 = \boldsymbol{x}^0 - (\nabla^2 f(\boldsymbol{x}^0))^{-1}\nabla f(\boldsymbol{x}^0) = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{b}$ (2pts). Note that f is a quadratic function and the Hessian is invertible (PD). Newton's method will find a stationary point in one iteration, and hence $\boldsymbol{x}^1$ is already a stationary point (actually optimal solution due to convexity) (2pts).

b) No (1pts). Let us choose $i = 1$ without loss of generality. Note that $\nabla f_1(\boldsymbol{x}) = 2(\boldsymbol{a}_1^\top\boldsymbol{x} - b_1)\boldsymbol{a}_1$. We will show by example that $-\nabla f_1(\boldsymbol{0})^\top\nabla f(\boldsymbol{0}) < 0$ is not always true (1pts). Let us construct

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ -\boldsymbol{a}_1^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

That is, $\boldsymbol{A}$ has $\boldsymbol{a}_1^\top$ and $-\boldsymbol{a}_1^\top$ as its first two rows while all remaining rows are all-zero vectors and $\boldsymbol{b}$ has the same first two entries while other entries are zeros. In this case, we have

$$-\nabla f_1(\boldsymbol{0})^\top\nabla f(\boldsymbol{0}) = -4b_1\boldsymbol{a}_1^\top\boldsymbol{A}^\top\boldsymbol{b} = 0,$$

which shows that $-\nabla f_1(\boldsymbol{0})$ is not a descent direction (3pts).

c) Note that $\boldsymbol{d}^k = -\nabla f(\boldsymbol{x}^k)$ is the search direction in the gradient descent method. Since f is quadratic, its second-order Taylor expansion is exact

$$\begin{aligned} f(\boldsymbol{x}^k + \alpha_k\boldsymbol{d}^k) - f(\boldsymbol{x}^k) &= \alpha_k\nabla f(\boldsymbol{x}^k)^\top\boldsymbol{d}^k + \frac{1}{2}\alpha_k^2\boldsymbol{d}^{k\top}\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{d}^k \quad \text{(2pts)} \\ &= -\alpha_k\|\nabla f(\boldsymbol{x}^k)\|^2 + \frac{1}{2}\alpha_k^2\nabla f(\boldsymbol{x}^k)^\top\nabla^2 f(\boldsymbol{x}^k)\nabla f(\boldsymbol{x}^k) \\ &= -8\alpha_k + 5\alpha_k^2. \quad \text{(2pts)} \end{aligned}$$

In order to let $\alpha_k = 1$ to be selected by backtracking line search, we need to satisfy the Armijo condition, thus

$$-8\alpha_k + 5\alpha_k^2 \le \gamma\alpha_k\nabla f(\boldsymbol{x}^k)^\top\boldsymbol{d}^k = -8\gamma\alpha_k, \quad \text{with } \alpha_k = 1. \quad \text{(2pts)}$$

This gives $(0 <)\gamma \le \frac{3}{8}$ (1pts).

Problem 5 (Integer Programming Modeling): (*12* points)

A company must produce at least 2000 units of a certain part. They can use one or more of the three production lines they own. For each production line, if one chooses to use it, then one has to produce at least 500 parts on that production line. The table below gives the relevant cost and capacity data.

Production line	Setup Cost	Production Unit Cost	Capacity (units)
1	600	2	800
2	100	10	1500
3	300	5	1200

The company wants to decide which production lines to use and how many parts to make on each production line. The objective is to minimize the total cost. Formulate this as an integer program.

Solution : Let x_i denote the number of units to produce on production line i and y_i denote whether we use production line i or not ($y_i = 1$ if we use production line i and $y_i = 0$ if we don't use production line i). (2pts)

The integer programming formulation for this problem is:

$$\begin{array}{lcr} \text{minimize} & 600y_1 + 100y_2 + 300y_3 + 2x_1 + 10x_2 + 5x_3 & \text{(3pts)} \\ \text{subject to} & x_1 \geq 500y_1 & \\ & x_2 \geq 500y_2 & \\ & x_3 \geq 500y_3 & \text{(3pts)} \\ & x_1 \leq 800y_1 & \\ & x_2 \leq 1500y_2 & \\ & x_3 \leq 1200y_3 & \text{(3pts)} \\ & x_1 + x_2 + x_3 \geq 2000 & \text{(1pt)} \\ & x_1, x_2, x_3 \in \mathbb{Z}^+, y_1, y_2, y_3 \in \{0, 1\} & \end{array}$$

Problem 6 (Branch-and-Bound Algorithm): (*20* points)

Consider the following knapsack problem:

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

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

Hint: It is easy to find the optimal solution to the LP relaxation of such problems.

In particular, one first ranks the value-weight ratio of all items. In this case, $13/5 > 7/3 > 9/4 > 3/2$. Then one sets the maximal value for the variables in the LP relaxation according to the value-weight ratio order. For example, the optimal solution to the LP relaxation for the initial problem is $x_1 = 1$ (that is the maximum one can set for x_1), $x_2 = 1$ (that is the maximum one can set for x_2) and $x_3 = 1/2$ (that is the maximum one can set for x_3 given x_1 and x_2 are set) and $x_4 = 0$. Also, in this problem, it is suggested to consider the smaller branch first (the branch in which the variable takes the small value).

Solution : We first solve the LP relaxation of the initial problem and get the optimal solution is $(1, 1, 1/2, 0)$ with objective value 24.5 (2pts).

Because this is not an integer solution, we define two branches (2pts).

a) (S1) Adding constraint $x_3 = 0$;

b) (S2) Adding constraint $x_3 = 1$.

Solving the LP relaxation of (S1), we get the optimal solution is $(1, 1, 0, 1)$ with objective value 23 (2pts). This is an integer solution so we don't need to further branch here. And 23 will be a lower bound for the entire problem. (2pts)

Then we solve the LP relaxation of (S2), we find the optimal solution to the LP relaxation is $(1, 1/3, 1, 0)$ with objective value 24.33 (2pts).

This is not an integer solution, so we further branch it into (2pts)

a) (S3) Adding constraint $x_2 = 0$;

b) (S4) Adding constraint $x_2 = 1$.

Solving the LP relaxation of (S3), we get the optimal solution is $(1, 0, 1, 1/2)$ with objective value 23.5 (2pts). However, this means that there can not be an integer solution with objective value more than 23 from this branch. Therefore, we abandon this branch (2pts).

Solving the LP relaxation of (S4), we get the optimal solution is $(0.6, 1, 1, 0)$ with objective value 23.8. For the same reason as in the above, we abandon this branch (2pts).

Therefore, $(1, 1, 0, 1)$ with objective value 23 is the optimal solution. (4pts)

A branch-and-bound tree of this problem is shown below: