FINAL EXAMINATION

January 2023

Question	Points	Score
True or False	15	
Nonlinear Optimization Algorithm	19	
KKT Conditions	19	
Integer Programming Formulation	15	
Convexity	16	
Optimality Condition for a Nonlinear Problem	16	
Total:	100	

- Please write down your **student ID** on the **answer paper**.
- Please justify your answers except Question 1.
- The exam time is 90 minutes.
- Two pieces of A4 paper with handwritten notes allowed (double-sided).
- No electronic devices including calculators are allowed.
- Even if you are not able to answer all parts of a question, write down the part you know. You will get corresponding credits to that part.

Question 1 [15 points]: True or False

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) [3 points] The set $\{x \in \mathbb{R}^n \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbb{R}^n$ with S_1 convex, is convex (Hint: for $x \in \mathbb{R}^n$ and $S \subseteq \mathbb{R}^n$, $x + S := \{x + y | y \in S\}$.).

Solution: True. (Convex set). $x + S_2 \subseteq S_1$ if $x + y \in S_1$ for all $y \in S_2$. Therefore

$$\{x \mid x + S_2 \subseteq S_1\} = \bigcap_{y \in S_2} \{x \mid x + y \in S_1\} = \bigcap_{y \in S_2} (S_1 - y),$$

the intersection of convex sets $S_1 - y$.

(b) [3 points] In the exact line search method, the new descent direction is orthogonal to the previous one.

Solution: True (Algorithm).

(c) [3 points] Consider the integer programming problem: $\min \mathbf{c}^\top \mathbf{x}$, subject to $\mathbf{Ax} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \mathbf{x}$ integer, where the entries of $\mathbf{A}, \mathbf{c}, \mathbf{b}$ are integer. Assume that integer programming problem is feasible. If the linear programming relaxation has optimal cost equal to $-\infty$, then the integer programming problem has optimal cost $-\infty$ as well.

Solution: True. (Integer Programming) let us suppose there is some integer solution x, as in the question. Since the relaxed program is unbounded, and since all coefficients A and b are integer, there is a direction

$$\delta \in \mathbb{Q}^n$$

of unboundedness described by rational coefficients, because such direction can be constructed from the final simplex table that shows the unboundedness, and this table will only contain rational values. More specifically, for every $\lambda \in \mathbb{R}_{\geq 0}$, we have that

$$A(x + \lambda\delta) \geq b$$

Because we assumed that the relaxed program has solution $-\infty$, we also know that $c^T\delta < 0$, since the objective function must go to $-\infty$ as λ goes to $+\infty$. However, because δ contains only rational values, the set

$$\{x + \lambda\delta : \lambda \in \mathbb{R}_{\geq 0}\}$$

must contain infinitely many integer points, for instance those of the form

$$\{x + n\lambda^*\delta : n \in \mathbb{N}\}$$

where λ^* is the smallest integer such that $\lambda^*\delta \in \mathbb{Z}^n$. Such λ^* exists because δ has only rational values.

(d) [3 points] If $C \subseteq \mathbb{R}^n$ is nonempty, closed and convex, and the norm $\|\cdot\|$ is strictly convex, then for every $x_0 \in \mathbb{R}^n$ there is exactly one $x \in C$ closest to x_0 (Hint: a function $f : \mathbb{R}^n \to \mathbb{R}$ is called strictly convex if and only if for all real $0 < t < 1$ and all $x_1, x_2 \in \mathbb{R}^n$ such that $x_1 \neq x_2$: $f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$.).

Solution: True. (Projection) There is at least one projection (this is true for any norm). And uniqueness can be proved. If the norm is strictly convex, suppose $u, v \in C$ with $u \neq v$ and $\|u - x_0\| = \|v - x_0\| = D$. Then $(1/2)(u+v) \in C$ and

$$\begin{aligned}\|(1/2)(u+v) - x_0\| &= \|(1/2)\,(u - x_0) + (1/2)\,(v - x_0)\| \\ &< (1/2)\,\|u - x_0\| + (1/2)\,\|v - x_0\| \\ &= D,\end{aligned}$$

so u and v are not the projection of x_0 on C.

(e) [3 points] In a constrained minimization problem, a point that satisfies KKT conditions is a local minimum.

Solution: False. (KKT) It can be a local maximum or a saddle point.

Question 2 [19 points]: Nonlinear Optimization Algorithm

Consider the following function:

$$f(x_1, x_2) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & \text{if } |x_2| \leq x_1 \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{otherwise} \end{cases} \tag{1}$$

Let $\gamma > 1$. Consider the gradient descent algorithm applied to f, with starting point $x^{(0)} = (\gamma, 1)^\top$ and an exact line search. Denote the iterate at k-th step as $x^{(k)} = (x_1^{(k)}, x_2^{(k)})^\top$.

(a) [9 points] Given $x^{(0)} = (\gamma, 1)^\top$, find the stepsize α_0 in exact line search.

Solution: Suppose $|x_2| \leq x_1$. The gradient of f at x is $(\frac{x_1}{\sqrt{x_1^2+\gamma x_2^2}}, \frac{\gamma x_2}{\sqrt{x_1^2+\gamma x_2^2}})^\top$. And

$$x - \alpha \nabla f(x) = \left(x_1 - \frac{\alpha x_1}{\sqrt{x_1^2 + \gamma x_2^2}}, x_2 - \frac{\alpha \gamma x_2}{\sqrt{x_1^2 + \gamma x_2^2}} \right)^\top.$$

Particularly, for $x^{(0)} = (\gamma, 1)^\top$, we have

$$x^{(0)} - \alpha \nabla f(x^{(0)}) = \left(\gamma - \frac{\alpha\gamma}{\sqrt{\gamma^2 + \gamma}}, 1 - \frac{\alpha\gamma}{\sqrt{\gamma^2+\gamma}} \right)^\top.$$

If $\alpha \leq \frac{1}{2}(\gamma+1)\sqrt{\gamma^2+\gamma}/\gamma$, we have $|1 - \frac{\alpha\gamma}{\sqrt{\gamma^2+\gamma}}| \leq \gamma - \frac{\alpha\gamma}{\sqrt{\gamma^2+\gamma}}$, and

$$\begin{aligned} f(x^{(0)} - \alpha\nabla f(x^{(0)})) =& \sqrt{\gamma^2 - \frac{2\alpha\gamma^2}{\sqrt{\gamma^2+\gamma}} + \frac{\alpha^2\gamma^2}{\gamma^2+\gamma} + \gamma - \frac{2\alpha\gamma^2}{\sqrt{\gamma^2+\gamma}} + \frac{\alpha^2\gamma^3}{\gamma^2+\gamma}} \\ =& \sqrt{\gamma^2 + \gamma - \frac{4\alpha\gamma^2}{\sqrt{\gamma^2+\gamma}} + \alpha^2\gamma} \end{aligned}$$

This is minimized by

$$\alpha_0 = \frac{2\gamma}{\sqrt{\gamma^2+\gamma}}.$$

If instead $\alpha > \frac{1}{2}(\gamma+1)\sqrt{(\gamma+1)/\gamma}$, then

$$f(x^{(0)} - \alpha\nabla f(x^{(0)})) \propto \alpha.$$

We conclude that $\alpha_0 = \frac{2\gamma}{\sqrt{\gamma^2+\gamma}}$ is optimal.

(b) [4 points] Derive the closed-form expression for $x^{(1)} = (x_1^{(1)}, x_2^{(1)})^\top$.

Solution:

$$x^{(1)} = x^{(0)} - \alpha \nabla f(x^{(0)}) = \left(\gamma - \frac{2\gamma}{\gamma+1}, 1 - \frac{2\gamma}{\gamma+1}\right) = \left(\gamma\frac{\gamma-1}{\gamma+1}, -\frac{\gamma-1}{\gamma+1}\right).$$

(c) [6 points] Suppose the closed-form expression for $x^{(k)} = (x_1^{(k)}, x_2^{(k)})^\top$ ($k \geq 0$) is $x_1^{(k)} = a_1 b_1^k$ and $x_2^{(k)} = b_2^k$ for some $b_1, b_2 \in (-1, 1)$. What is the convergent point of $x^{(k)}$? Is it the optimal solution to $\min_{x\in\mathbb{R}^2} f(x_1, x_2)$?

Solution:
Clearly $x^{(k)}$ converges to $(0, 0)^\top$. It is not optimal since the problem is unbounded.

Question 3 [19 points]: KKT Conditions

Consider the problem:

$$\begin{array}{ll} \text{minimize} & -2\left(x_1-2\right)^2 - x_2^2 \\ \text{subject to} & x_1^2 + x_2^2 \leq 25 \\ & x_1 \geq 0 \end{array}$$

(a) [6 points] Write down the KKT optimality condition.

Solution:

$$L(x_1, x_2, \lambda_1, \lambda_2) = -2(x_1-2)^2 - x_2^2 + \lambda_1(x_1^2 + x_2^2 - 25) - \lambda_2 x_1$$

The KKT condition is:
Main condition:

$$-4\left(x_1-2\right) + 2\lambda_1 x_1 - \lambda_2 = 0, \quad -2x_2 + 2\lambda_1 x_2 = 0,$$

Dual feasibility:

$$\lambda_i \geqslant 0,$$

Complementarity:

$$\lambda_1(x_2^2 + x_2^2 - 25) = 0, \quad \lambda_2 x_1 = 0.$$

Primal feasibility:

$$x_1^2 + x_2^2 \leq 25, \quad x_1 \geq 0.$$

(b) [8 points] Find all the KKT-points and the corresponding multiplier vectors.

Solution: Consider the main condition: $-2x_2 + 2\lambda_1 x_2 = 0$,
Case1: When $x_2 = 0$, consider the complementarity condition $\lambda_1(x_1^2 + x_2^2 - 25) = 0$

i. Suppose $\lambda_1 = 0$, consider the complementarity condition $\lambda_2 x_1 = 0$,

(a) if $\lambda_2 = 0$, the main condition $\frac{\partial L}{\partial x_1} = 0$ becomes $-4x_1 + 8 = 0$, then we have KKT pairs: $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (2, 0, 0, 0)$;

(b) if $x_1 = 0$, the main condition $\frac{\partial L}{\partial x_1} = 0$ becomes $8 - \lambda_2 = 0$, then we have KKT pairs: $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (0, 0, 0, 8)$.

ii. Suppose $x_1^2 + x_2^2 - 25 = 0$, then we have $x_1 = 5$, $\lambda_2 = 0$ since $\lambda_2 x_1 = 0$. The main condition $\frac{\partial L}{\partial x_1} = 0$ becomes $-12 + 10\lambda_1$, then we have KKT pairs: $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (5, 0, \frac{6}{5}, 0)$.

Case2: When $\lambda_1 = 1$, the complementarity condition $\lambda_1(x_1^2+x_2^2-25) = 0$ becomes $x_1^2+x_2^2-25 = 0$. Consider $\lambda_2 x_1 = 0$,

i. suppose $x_1 = 0$, then $x_2 = 5$ or $x_2 = -5$. The main condition $\frac{\partial L}{\partial x_1} = 0$ becomes $8 - \lambda_2 = 0$, then we have KKT pairs: $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (0, 5, 1, 8)$. and $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (0, -5, 1, 8)$.

ii. suppose $\lambda_2 = 0$, the main condition $\frac{\partial L}{\partial x_1} = 0$ becomes $-2x_1 + 8 = 0$, we have $x_1 = 4$,$x_2 = 3$ or -3. We have KKT pairs: $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (4, 3, 1, 0)$. and $(x^*, \lambda^*) = (x_1, x_2, \lambda_1, \lambda_2) = (4, -3, 1, 0)$.

(c) [5 points] Find the global minimizer(s) of the problem.

Solution:

i. when $(x_1, x_2) = (2, 0)$, the objective value is -4;

ii. when $(x_1, x_2) = (0, 0)$, the objective value is -8;

iii. when $(x_1, x_2) = (5, 0)$, the objective value is -18;

iv. when $(x_1, x_2) = (0, 5)$ or $(x_1, x_2) = (0, -5)$, the objective value is -33;

v. when $(x_1, x_2) = (4, 3)$ or $(x_1, x_2) = (4, -3)$, the objective value is -17;

The global minima are $(0, 5)$ and $(0, -5)$, with objective value equals to -33.

Question 4 [15 points]: Integer Programming Formulation

(a) [5 points] Truck 1 can transfer four machines A, B, C, D to the market. The profit of each machine is 300 dollars(A), 500 dollars(B), 400 dollars(C) and 550 dollars(D), respectively. And the weight of each machine is 3.5 tons(A), 5 tons(B), 4.5 tons(C) and 2 tons(D). The capacity of this truck is 10 tons. Please give an integer programming formulation to choose the machines that maximizes the profits of the truck. The truck can only transfer at most three machines. If machine D is chosen in this truck, machine A must be chosen as well.

Solution: We can express the problem as follows, x_i is binary variable, $x_i = 1$ if machine i is chosen. p_i is the profit of machine i, w_i is the weight of machine i.

$$\begin{array}{ll} \text{maximize}_{\text{x}} & \sum_{i=1}^{4} x_i * p_i \\ \text{subject to} & \sum_{i=1}^{4} x_i * w_i \leq 10 \\ & \sum_{i=1}^{4} x_i \leq 3 \\ & x_1 \geq x_4 \\ & x_i \in \{0, 1\}, \forall i = 1....4 \end{array}$$

(b) [10 points] Use the branch-and-bound method to solve the following integer program. Please specify the branch-and-bound tree and what you did at each node.

$$\begin{aligned} \text{maximize} \quad & 4x_1 + 9x_2 \\ \text{subject to} \quad & 2x_1 + 5x_2 \leq 22 \\ & x_1 \leq 5 \\ & x_2 \leq 3 \\ & x_1, x_2 \geq 0 \\ & x_1, x_2 \in \mathbb{Z}. \end{aligned}$$

(Hint: The feasible set is a polyhedron in the two-dimensional space with five extreme points.)

Solution:
We first draw the feasible set

$$\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : 2x_1 + 5x_2 \leq 22, x_1 \leq 5, x_2 \leq 3, x_1 \geq 0, x_2 \geq 0\}$$

of the first LP relaxation.
The five extreme points of the polyhedron Ω are $p_1 = (0,0)^T, p_2 = (5,0)^T, p_3 = (5,2.4)^T, p_4 = (3.5,3)^T, p_5 = (0,3)^T$. The optimal solution of the relaxed LP is attained at p_3 with optimal value 41.6, which means that the optimal function value of the integer program needs to be less or equal than 41.6.
We branch on $x_2 = 2.4$. We consider the two branches:
• (S1): $x_2 \leq 2$.
• (S2): $x_2 \geq 3$.
For (S1), the optimal solution of the LP relaxation is given by $(5,2)^T$ with objective value 38. This is an integer solution and we obtain the lower bound 38.
For (S2), using the sketch, the optimal solution satisfies $x_2 = 3$ and $2x_1 + 5x_2 = 22$. This gives $(3.5,3)^T$ with value 41.
We need to further branch on x_1. We consider the two branches:
• (S3): $x_1 \leq 3$.
• (S4): $x_1 \geq 4$.
We immediately see that (S4) is infeasible. For (S3), the extreme point with $x_1 = 3$ and $2x_1+5x_2 = 22$, together with $x_2 \leq 3$ needs to be the solution. This yields $(3,3)^T$ and the value is 39, which is larger than 38.
Therefore, the optimal solution for this integer program is $(3,3)^T$ with objective value 39.

Question 5 [16 points]: Convexity

Let x be a real-valued random variable with $Pr(x = a_i) = p_i, i = 1, \ldots, n$, where $a_1 < a_2 < \cdots < a_n$ and $Pr(x = a_i) = p_i$ indicates that the probability of $x = a_i$ is equal to p_i. Each probability distribution $p \in \mathbb{R}^n$ lies in the standard probability simplex $P = \{p | \sum_{i=1}^n p_i = 1, p_i \geq 0 \text{ for all } i\}$. For which of the following conditions is the set of $p \in P$ that satisfies the condition convex? If the set of $p \in P$ that satisfies the condition is a convex set, please provide a proof or explanation; otherwise, please provide counterexamples.

(a) [4 points] For a given function $f : \mathbb{R} \to \mathbb{R}$. $\alpha \leq \mathbb{E}f(x) \leq \beta$, where $\mathbb{E}f(x)$ is the expected value of $f(x)$, i.e., $\mathbb{E}f(x) = \sum_{i=1}^n p_i f(a_i)$.

Solution: Yes. $\mathbb{E}f(x) = \sum_{i=1}^n p_i f(a_i)$, so the constraint is equivalent to two linear inequalities

$$\alpha \leq \sum_{i=1}^n p_i f(a_i) \leq \beta.$$

(b) [4 points] $\mathbb{E}x^2 \geq \alpha$ for a given positive number α.

Solution: Yes. The constraint is equivalent to a linear inequality

$$\sum_{i=1}^{n} p_i a_i^2 \geq \alpha.$$

(c) [4 points] $Var(x) \leq \alpha$ for a given positive number α, where $Var(x) = \mathbb{E}(x - \mathbb{E}x)^2 = \mathbb{E}x^2 - (\mathbb{E}x)^2$ is the variance of x.

Solution: No. The constraint

$$Var(x) = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{i=1}^{n} p_i a_i\right)^2 \leq \alpha$$

is not convex in general. As a counterexample, we can take $n = 2, a_1 = 0, a_2 = 1$, and $\alpha = 1/5$. $p = (1, 0)$ and $p = (0, 1)$ are two points that satisfy $Var(x) \leq \alpha$, but the convex combination $p = (1/2, 1/2)$ does not.

(d) [4 points] $Var(x) \geq \alpha$ for a given positive number α.

Solution: Yes. This constraint is equivalent to

$$Var(x) = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{i=1}^{n} p_i a_i\right)^2 \geq \alpha$$

$$p^T A p - b^T p + \alpha \leq 0$$

where $b_i = a_i^2$ and $A = aa^T$. This defines a convex set, since the matrix aa^T is positive semidefinite.

Question 6 [16 points]: Optimality Condition for a Nonlinear Problem

Consider the following optimization problem.

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega, \end{aligned} \tag{2}$$

where $\Omega \subset \mathbb{R}^n$ is a non-empty, closed and convex set and $f \in \mathcal{C}^2(\mathbb{R})$, i.e., f is continuous, and both the gradient and the Hessian matrix of f are continuous mappings on $\mathbb{R}^n$.

(a) [8 points] Given $x^* \in \Omega$, suppose that there exists $c > 0$ such that for all feasible directions d at x^*, (i) $d^\top \nabla f(x^*) \geq 0$ and (ii) $d^\top \nabla f(x^*)\, d \geq c\|d\|^2$. Show that x^* is a strict local minimizer of f over Ω. (Hint: Define $B_{\varepsilon(y)} := \{x \in \mathbb{R}^n : \|x - y\| \leq \varepsilon\}$ to be an open ball in $\mathbb{R}^n$ centered at y with radius $\varepsilon > 0$. We call x^* a strict local minimizer of (2) if $x^* \in \Omega$ and there is $\varepsilon > 0$ with $f(x) > f(x^*)$ for all $x \in (\Omega \cap B_{\varepsilon(x^*)}) \backslash \{x^*\}$.)

Solution: Since $f \in \mathcal{C}^2(\mathbb{R}^n)$, consider the Second-order Taylor extension at x^*.

$$f(x^* + d) = f(x^*) + d^T \nabla f(x^*) + d^\top \nabla f(x^*)\, d + o(\|d\|^2)$$

By the definition of $o(\|d\|)$, there exists $\delta > 0$, for any $\|d\| \leq \delta$, we have $o(\|d\|^2) \geq -\frac{c}{2}\|d\|^2$. Thus, for any d satisfied that $x^* + d \in O_\delta(x^*) \cap S$ and $d \neq 0$ (where $O_\delta(x^*)$ is an open ball centered at x^*, which means for all $x \in O_\delta(x^*)$, $\|x - x^*\| \leq \delta$), we have

$$f(x^* + d) \geq f(x^*) + c\|d\|^2 - \frac{c}{2}\|d\|^2 > f(x^*)$$

Therefore, x^* is a strictly local minimizer.

(b) [8 points] Suppose that $z \in \Omega$ satisfies the first order optimality condition:

$$\forall x \in \Omega, \quad \langle \nabla f(z), x - z \rangle \geq 0,$$

and the strict version of second-order optimality condition:

$$\forall x \neq z \in \Omega, \quad \left\langle \nabla^2 f(z)(x - z), x - z \right\rangle > 0,$$

Prove or disprove, z is a local minimum of f on Ω.

Solution: A counter-example is shown below.

$$\begin{aligned} \min \quad & f(x_1, x_2) = \frac{1}{2}x_1^2 - x_1 - x_2^3 \\ \text{s.t.} \quad & x \in S := \{x_1^2 + x_2^2 \leq 1\} \end{aligned}$$

Set $z = (1, 0)^T$, we have $\nabla f(z) = (0, 0)^T$, $\nabla^2 f(z) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
Thus, $\langle \nabla f(z), x - z \rangle = 0$ and $\langle \nabla^2 f(z)(x - z), x - z \rangle > 0$.
However, take $z_n = (1 - \frac{2}{3n+3}, \frac{1}{\sqrt{n+1}})$. $z_n \to z$.
$(1 - \frac{2}{3n+3})^2 + (\frac{1}{\sqrt{n+1}})^2 = 1 - \frac{1}{3n+3} + \frac{4}{9(n+1)^2} \leq 1$ which means z_n is feasible.
But $f(z_n) = -\frac{1}{2} + \frac{2}{9(n+1)^2} - \frac{1}{(n+1)\sqrt{n+1}} < -\frac{1}{2} = f(z)$ for all z_n.