**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 (Optimality Conditions):** (*20* points)

Consider the constrained optimization problem

$$\text{minimize}_{\boldsymbol{x}\in\mathbb{R}^2}\ f(\boldsymbol{x}) := \frac{1}{4}x_1^4 + (x_1 - x_2)x_1^2 + \frac{1}{2}x_2^2 - x_1^2 \quad \text{subject to} \quad x_1 \ge 0, \quad x_2 \ge x_1^2. \tag{1}$$

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

b) Write down the KKT conditions for (1). Compute and find all corresponding KKT points of the minimization problem (1).

c) Prove that the point $\boldsymbol{x}^* = (1,1)^\top$ is a strict local minimum of the problem (1).

***Solution :***

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

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} x_1^3 + 3x_1^2 - 2x_1x_2 - 2x_1 \\ -x_1^2 + x_2 \end{pmatrix}, \quad \nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 3x_1^2 + 6x_1 - 2x_2 - 2 & -2x_1 \\ -2x_1 & 1 \end{pmatrix}.$$

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) Let us define $g_1(\boldsymbol{x}) := -x_1$ and $g_2(\boldsymbol{x}) := x_1^2 - x_2$. The KKT conditions for (1) are then given by:

$$\nabla f(\boldsymbol{x}) + \nabla g(\boldsymbol{x}) \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} x_1^3 + 3x_1^2 - 2x_1x_2 - 2x_1 \\ -x_1^2 + x_2 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \end{pmatrix} \lambda_1 + \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix} \lambda_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$g_1(\boldsymbol{x}), g_2(\boldsymbol{x}) \le 0$, $\lambda_1, \lambda_2 \ge 0$, and $\lambda_1 g_1(\boldsymbol{x}) = \lambda_2 g_2(\boldsymbol{x}) = 0$.

To find all KKT points, we consider several cases.

**Case 1:** $\lambda_1 > 0$. By the complementarity conditions, this implies $x_1 = 0$ and the main condition simplifies to $-\lambda_1 = 0$ and $x_2 - \lambda_2 = 0$. This is a contradiction.

**Case 2:** $\lambda_2 > 0$. We then have $x_2 = x_1^2$ and the main condition reduces to

$$-x_1^3 + 3x_1^2 - 2x_1 - \lambda_1 + 2x_1\lambda_2 = 0 \quad \text{and} \quad -\lambda_2 = 0.$$

This again leads to a contradiction.

**Case 3:** $\lambda_1 = \lambda_2 = 0$. In this case, we essentially need to solve the equation $\nabla f(\boldsymbol{x}) = \mathbf{0}$. By the second equation, we have $x_2 = x_1^2$ and we obtain $-x_1^3+3x_1^2-2x_1 = -x_1(x_1^2-3x_1+2) = 0$. This equation holds for $x_1 = 0$ and $x_{1,2/3} = (3 \pm \sqrt{9-4\cdot 2})/2 = (3\pm 1)/2 = 2\,/\,1$. Hence, the function $f$ has the three stationary points:

$$\boldsymbol{x}_1^* = (0,0)^\top, \quad \boldsymbol{x}_2^* = (2,4)^\top, \quad \boldsymbol{x}_3^* = (1,1)^\top.$$

As all of these points are feasible ($x_1 \geq 0$ and $x_2 = x_1^2$), $\boldsymbol{x}_1^*$, $\boldsymbol{x}_2^*$, and $\boldsymbol{x}_3^*$ are exactly the KKT points of (1).

13 pts in total. 3 pts for complete and correct KKT conditions (-0.5 for missing parts / errors). 1pt for $\nabla g_1$ and $\nabla g_2$. 2 pts for Case 1 ($\lambda_1 > 0$). 2 pts for Case 2 ($\lambda_2 > 0$). 5 pts for Case 3 (1pt for $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$, 1pt each for $\boldsymbol{x}_i^*$; 1pt for noting feasibility).

c) Since $\boldsymbol{x}^*$ is a stationary point of $f$, we first check the Hessian $\nabla^2 f$ at $\boldsymbol{x}^*$:

$$\nabla^2 f(\boldsymbol{x}^*) = \begin{pmatrix} 3+6-2-2 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}.$$

It holds that $\operatorname{tr}(\nabla^2 f(\boldsymbol{x}^*)) = 6$ and $\det(\nabla^2 f(\boldsymbol{x}^*)) = 5 - 4 = 1$. Consequently, $\nabla^2 f(\boldsymbol{x}^*)$ is positive definite and $\boldsymbol{x}^*$ is a strict local minimum of $f$ (in the unconstrained sense). But then it also needs to be a strict local minimizer of the constrained problem (1).

4 pts in total. 1 pt for $\nabla^2 f$. 1 pt for checking and verifying definiteness of $\nabla^2 f$. 1 pt for $\boldsymbol{x}^* =$ local minimum of $f$. 1pt for $\boldsymbol{x}^* =$ also local minimum of (1).

---

**Problem 2 (Convex Functions):** (*16* points)

Investigate whether each of the following functions is convex, concave, or neither convex nor concave. Explain your answer!

a) $f : \mathbb{R}^2 \to \mathbb{R}$ and $f(\boldsymbol{x}) = (1 - x_1 x_2)^2$.

b) $f : X \to \mathbb{R}$, where $X = \mathbb{R}^n_{++} := \{\boldsymbol{x} \in \mathbb{R}^n : x_i > 0, \text{ for all } i\}$ and $f(\boldsymbol{x}) = \sum_{i=1}^n x_i \log(x_i)$.

c) $f : X \to \mathbb{R}$, where $X = \mathbb{R}^2_{++} = \{(x_1, x_2) \in \mathbb{R}^2 : x_i > 0, \ i = 1, 2\}$ and $f(\boldsymbol{x}) = -\log(1 + x_1 + x_2)$.

d) $f : X \to \mathbb{R}$, where $X = \mathbb{R}^2_{++}$ and $f(\boldsymbol{x}) = \frac{x_1}{x_2^2}$.

***Solution :***

a) (4 points.) $f$ is not convex. Consider $\boldsymbol{x} = (1, 1)^\top$, $\boldsymbol{y} = (-1, -1)^\top$, and $\lambda = 0.5$. Then $f(\boldsymbol{x}) = f(\boldsymbol{y}) = 0$ and $f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) = f(\boldsymbol{0}) = 1$. Therefore, we have

$$\lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}) < f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}).$$

b) (4 points.) Convex. The set $\mathbb{R}^n_{++}$ is obviously convex. We have

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j} [\log(x_i) + 1] = \begin{cases} \frac{1}{x_i}, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Hence, $\nabla^2 f(\boldsymbol{x})$ is a diagonal matrix with all entries being positive, and therefore it is positive semi-definite for all $\boldsymbol{x} \in X$.

c) (4 points.) Convex. We have

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{1}{(1+x_1+x_2)^2} & \frac{1}{(1+x_1+x_2)^2} \\ \frac{1}{(1+x_1+x_2)^2} & \frac{1}{(1+x_1+x_2)^2} \end{bmatrix} = \frac{1}{(1 + x_1 + x_2)^2} \mathbb{1}\mathbb{1}^\top.$$

Therefore, $\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq 0$ for all $\boldsymbol{v}$, which means it is positive semi-definite. (Or you can check the positive semi-definiteness by writing out the full expression for $\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{v}$.)

d) (4 points.) Not convex.

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 0 & -\frac{2}{x_2^3} \\ -\frac{2}{x_2^3} & \frac{6x_1}{x_2^4} \end{bmatrix}$$

This is a $2 \times 2$ matrix, with trace of $\frac{6x_1}{x_2^4} \geq 0$ and determinant of $-\frac{4}{x_2^6} < 0$. Therefore, there is one positive eigenvalue and one negative eigenvalue.

Rubric: Receive points only with suitable explanations.

---

**Problem 3 (Algorithms):** (*20* points)

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ where

$$f(x, y) := \frac{3}{4} x^{\frac{4}{3}} + \frac{1}{2}(x + y)^2.$$

a) Show that the function $f$ is convex. Find the global minimizer of the problem $\min_{x,y} \ f(x, y)$.

b) Given a point $(x^k, y^k)$ with $x^k \neq 0$ write out an explicit expression for $(x^{k+1}, y^{k+1})$ after one iteration of Newton's method applied to minimize this function.

Then, prove that starting from an initial point $(x^0, y^0)$ with $x^0 \neq 0$, Newton's method does not converge to the global minimizer.

**Hint:** The inverse of a $2 \times 2$ matrix can be simply computed via the formula:

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \to \quad \boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \quad a, b, c, d \in \mathbb{R}.$$

c) Starting from the initial point $(x^0, y^0) = (1, 1)$, compute the iterate $(x^1, y^1)$ after one iteration of the gradient descent method with a constant step size $\frac{1}{2}$ (without line-search).

d) Consider minimizing $f$ subject to the constraints $(x, y) \in X$ with $X = \{(x, y) : x + y = 1\}$. Starting from $(x^0, y^0) = (1, 1)$, compute $(x^1, y^1)$ after one iteration of the projected gradient descent method with a constant step size $\frac{1}{2}$ (without line-search).

***Solution :***

a) Define $f_1(x, y) = \frac{3}{4} x^{\frac{4}{3}}$ and $f_2(x, y) = \frac{1}{2}(x + y)^2$. Consider $g(x) = \frac{3}{4} x^{\frac{4}{3}}$. Now we first prove $g$ is convex (very similar to Problem 2d in Assignment 7): we can show $g$ is convex over $x > 0$ by checking the positivity of its second derivative. Assume both $x_1$ and $x_2$ are nonzero. For any $\lambda \in [0, 1]$

$$\begin{aligned} g(\lambda x_1 + (1 - \lambda) x_2) &= (\lambda x_1 + (1 - \lambda) x_2)^{4/3} \\ &= |\lambda x_1 + (1 - \lambda) x_2|^{4/3} \\ &\leq (|\lambda x_1| + |(1 - \lambda) x_2|)^{4/3} \quad \text{(triangle inequality)} \\ &= (\lambda |x_1| + (1 - \lambda)|x_2|)^{4/3} \\ &\leq \lambda |x_1|^{4/3} + (1 - \lambda)|x_2|^{4/3} \quad \text{(convexity of } x^{4/3} \text{ over } x > 0) \\ &= \lambda x_1^{4/3} + (1 - \lambda) x_2^{4/3} \\ &= \lambda g(x_1) + (1 - \lambda) g(x_2). \end{aligned}$$

It remains to consider $x_1 = 0$ or $x_2 = 0$. Without loss of generality, assume that $x_2 = 0$. Then for all $x_1$ and $\lambda \in [0, 1]$

$$g(\lambda x_1 + (1-\lambda)x_2) = \lambda^{4/3} x_1^{4/3} \le \lambda x_1^{4/3} = \lambda g(x_1) + (1-\lambda) g(x_2).$$

Now we conclude that $g$ is convex. (3 points for showing convexity of $g$. We didn't check using Hessian, because it is not well defined at 0). Then it is easy to show that $f_1$ is convex. Obviously, the quadratic function $f_2$ is convex. Hence, $f = f_1 + f_2$ is convex. (2 points. Some simple explanation is enough. )

$(0, 0)$ is the unique global minimizer because $f(x, y) > 0$ for all $(x, y) \neq (0, 0)$. (2 points.)

b) The gradient of the function is

$$\nabla f(x, y) = \begin{bmatrix} x^{\frac{1}{3}} + x + y \\ x + y \end{bmatrix}.$$

The Hessian is

$$\nabla^2 f(x, y) = \begin{bmatrix} \frac{1}{3} x^{-\frac{2}{3}} + 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

(2 points for computations of gradient and Hessian.) When $x \neq 0$, the inverse of the Hessian is

$$\nabla^2 f(x, y)^{-1} = \begin{bmatrix} 3x^{\frac{2}{3}} & -3x^{\frac{2}{3}} \\ -3x^{\frac{2}{3}} & 1 + 3x^{\frac{2}{3}} \end{bmatrix}.$$

Then Newton's method has the update:

$$\begin{aligned} \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} &= \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 3x_k^{\frac{2}{3}} & -3x_k^{\frac{2}{3}} \\ -3x_k^{\frac{2}{3}} & 1 + 3x_k^{\frac{2}{3}} \end{bmatrix} \cdot \begin{bmatrix} x_k^{\frac{1}{3}} + x_k + y_k \\ x_k + y_k \end{bmatrix} \\ &= \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 3x_k \\ -2x_k + y_k \end{bmatrix} = \begin{bmatrix} -2x_k \\ 2x_k \end{bmatrix}. \end{aligned}$$

(3 points for the update. Instead of Hessian inverse, it can also be done by solving a 2-by-2 linear system. Give 1 point if the formula of Newton's method is provided but without correct computation.)

Starting from $(x_0, y_0)$, the next iterate is $(x_1, y_1) = (-2x_0, 2x_0)$. Further after $k$ iterations, we have $((-2)^k x_0, 2 \cdot (-2)^{k-1} x_0)$. Therefore, it will not converge to $(0, 0)$.

(2 points for explanations here.)

c) (2 points.) After one iteration of gradient descent, we have

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k \left( x_k^{\frac{1}{3}} + x_k + y_k \right) = (1 - \alpha_k) x_k - \alpha_k x_k^{\frac{1}{3}} - \alpha_k y_k, \\ y_{k+1} &= y_k - \alpha_k (x_k + y_k) = (1 - \alpha_k) y_k - \alpha_k x_k. \end{aligned}$$

With $(x_0, y_0) = (1, 1)$ and $\alpha_0 = 1/2$, we have $(x_1, y_1) = (-1/2, 0)$.

d) Projection of $(x, y)$ onto $X$ is $(\frac{x-y+1}{2}, \frac{y-x+1}{2})$, which can be computed with the formula from class (or using Lagrangian): $X = \{[x; y] : \boldsymbol{A}[x; y] = b\}$ with $\boldsymbol{A} = [1, 1]$ and $b = 1$ and the projection is

$$\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \left( \boldsymbol{A} \begin{bmatrix} x \\ y \end{bmatrix} - b \right)$$

(2 points for projection.)

Then the iteration of the projection gradient descent method is just projecting part c) to $X$:

$$x_{k+1} = \frac{x_k - \alpha_k x_k^{\frac{1}{3}} - y_k + 1}{2},$$
$$y_{k+1} = \frac{y_k + \alpha_k x_k^{\frac{1}{3}} - x_k + 1}{2}.$$

With $(x_0, y_0) = (1, 1)$ and $\alpha_k = 1/2$, we have $(x_1, y_1) = (1/4, 3/4)$.

(2 points.)

---

**Problem 4 (True or False):** (*12* points)

State whether each of the following statements is *True* or *False*. For each statement provide a short explanation, counterexample, or proof. Only answers with full explanations will be graded. (Short answers of the form "true" or "false" will not be accepted).

a) The derivative $f'$ of the function $f(x) = \sin(x)$ is Lipschitz continuous.

b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. The point $\boldsymbol{x}^*$ is a stationary point of $f$ if and only if $\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{h} = 0$ for all $\boldsymbol{h} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$.

c) We consider the standard integer program

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \boldsymbol{x} \in \mathbb{Z}^n. \end{aligned} \tag{2}$$

If the integer program (2) is feasible, i.e., the feasible set is nonempty, then the LP relaxation also must be feasible.

***Solution :***

a) **True.** It holds that $f'(x) = \cos(x)$ and $f''(x) = -\sin(x)$ and we have $|f''(x)| \leq 1$ for all $x \in \mathbb{R}$. Thus, $f'$ is Lipschitz continuous with $L = 1$. (4 pts)

b) **True.** If $\boldsymbol{x}^*$ is a stationary point then we have $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ and the condition is obviously satisfied. On the other hand, let us assume $\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{h} = 0$ for all $\boldsymbol{h} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$. In the case $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, there is nothing to show. Hence, let us assume $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$. But then setting $\boldsymbol{h} = \nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$ implies $\|\nabla f(\boldsymbol{x}^*)\|^2 = 0$ which is a contradiction to $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$ and thus, this case cannot occur. Together, we can infer $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ which finishes the proof. (4 pts)

c) **True.** The feasible set of the LP relaxation is larger and contains the integer points of the original problem. Hence, the LP relaxation cannot be infeasible. (4 pts)

---

**Problem 5 (Integer Programming Modeling):** (*12* points)
A small manufacturing company produces two types of wooden furniture: chairs and tables. The production process involves two main resources: labor ("carpentry hours") and materials (wood in "board feet"). The company wants to determine the optimal number of chairs and tables to produce per day to maximize the profit while respecting resource constraints.

| **Category** | **Details** |
|---|---|
| Profit per item: | |
| Chair | \$20 |
| Table | \$50 |
| Resources per item – Work hours: | |
| Chair | 2 hours |
| Table | 5 hours |
| Resources per item – Wood (board feet): | |
| Chair | 4 board feet |
| Table | 10 board feet |

| **Category** | **Details** |
|---|---|
| Daily limits: | |
| Total work hours | 40 hours |
| Total wood available | 100 board feet |
| Extra rules: | |
| Minimum chairs | At least 2 per day |
| Maximum tables | No more than 8 per day |

Formulate this problem as an integer programming (IP) problem by: 1.) Choosing the decision variables; 2.) Formulating the goal (objective); 3.) Listing all of the constraints. You are not required to solve this problem.

***Solution :*** Decision Variables: Let:

- $x_1$ = Number of chairs produced per day (integer).
- $x_2$ = Number of tables produced per day (integer).

(2 points.)

Objective Function: Maximize profit: Maximize $Z = 20x_1 + 50x_2$ (2 points.)

Constraints:

a) **Labor constraint:** $2x_1 + 5x_2 \leq 40$

b) **Materials constraint:** $4x_1 + 10x_2 \leq 100$

c) **Minimum chairs constraint:** $x_1 \geq 2$

d) **Maximum tables constraint:** $x_2 \leq 8$

e) **Non-negativity and integer constraints:** $x_1, x_2 \geq 0$ and integers.

(8 points.) a,b,e each 2 points, c+d=2points.

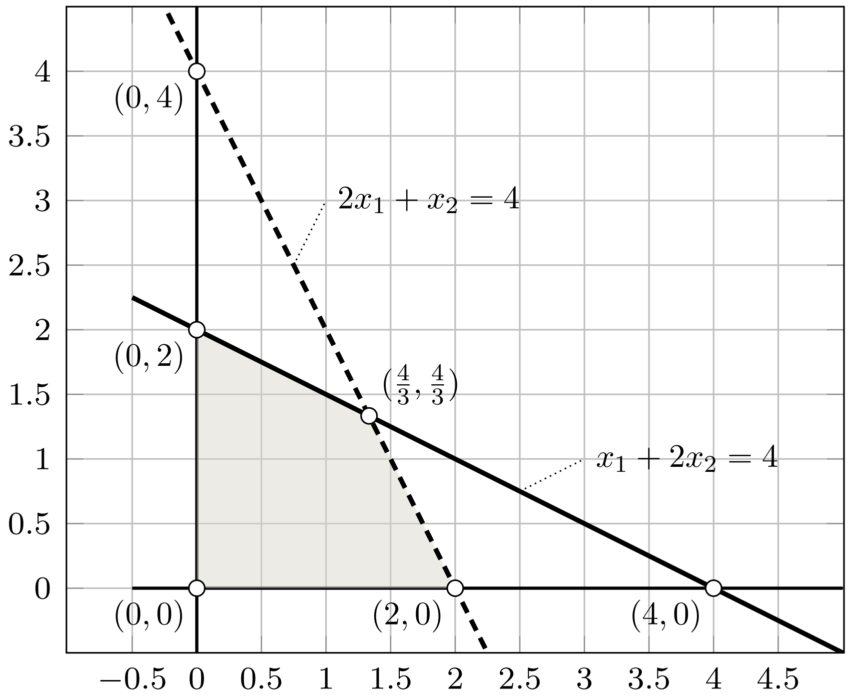
**Problem 6 (Branch-and-Bound):** (*20* points)

Consider the following integer programming problem:

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

Use the branch-and-bound algorithm to solve this problem by hand. Show the steps, including the branching and bounding process, and identify the optimal solution and optimal objective value.

**Hint:** In order to solve the LP relaxations you can use a graphical approach or check the corresponding extreme points. The feasible set is a polyhedron in the two-dimensional space. A visualization is shown below.




***Solution :*** First, solve the problem without the integer constraints.

S1: Linear Relaxation Problem:
Maximize $Z = 3x_1 + 2x_2$
Subject to:
$2x_1 + x_2 \leq 4$
$x_1 + 2x_2 \leq 4$
Solution: $(4/3, 4/3), 20/3$. Since $x_1$ and $x_2$ are not integers, we proceed with branching.
**S2:** $+x_1 \leq 1$
**S3:** $+x_1 \geq 2$
(5 points.)

S2: Maximize $Z = 3x_1 + 2x_2$
Subject to:

1. $2x_1 + x_2 \leq 4$
2. $x_1 + 2x_2 \leq 4$
3. $x_1 \leq 1$
4. $x_1, x_2 \geq 0$

Solution: $(1, 3/2)$, 6. This is not an integer, so we branch further on $x_2$.
(5 points.)

S4: $+x_2 \leq 1$
Maximize $Z = 3x_1 + 2x_2$
Subject to:
$2x_1 + x_2 \leq 4$
$x_1 + 2x_2 \leq 4$
$x_1 \leq 1$
$x_2 \leq 1$
$x_1, x_2 \geq 0$
Solution: $(1, 1), 5$. This is an integer, 5 is a lower bound.
(3 points.)

S5: $+x_2 \geq 2$
Maximize $Z = 3x_1 + 2x_2$
Subject to:
$2x_1 + x_2 \leq 4$
$x_1 + 2x_2 \leq 4$
$x_1 \leq 1$
$x_2 \geq 2$
$x_1, x_2 \geq 0$
Solution: $(0, 2), 4$. This is worse than the current best ($Z = 5$), so we discard it.
(3 points.)

S3: $x_1 \geq 2$
Maximize $Z = 3x_1 + 2x_2$
Subject to:
$2x_1 + x_2 \leq 4$
$x_1 + 2x_2 \leq 4$
$x_1 \geq 2$
$x_1, x_2 \geq 0$
Solution: $(2, 0), 6$. No further branching is needed because we cannot improve beyond the objective value of 6.
(3 points.)

Finally, compare $(1, 1)$ and $(2, 0)$'s objectives, we confirm that $(2, 0)$ is the optimal solution with optimal obj 6.
(1 point.)