SAMPLE FINAL 23 FALL

Sample

Question	Points	Score
True or False	15	
KKT Conditions	15	
Convexity	18	
Branch-and-Bound Method	20	
Algorithm	21	
Descent with Constant Step Size	11	
Total:	100	

- Please write down your **name** and **student ID** on the **answer paper**.
- Please justify your answers except Question 1.
- 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] Consider the function $f(x) = |x| + x$. The unconstrained minimization problem, i.e., $\min_{x\in\mathbb{R}} f(x)$, does not have a global minimizer.

Solution: False.

(b) [3 points] If the linear independence constraint qualification (LICQ) holds at a local minimizer x of a constrained optimization problem, then the associated Lagrangian multipliers λ (for inequality constraints) and μ (for equality constraints) must be unique.

Solution: True.

(c) [3 points] Let y be the Euclidean projection of x onto a convex, closed, nonempty set Ω. Then, y is the point in Ω that is closest to x, i.e., $\|y - x\| < \|z - x\|, \forall z \in \Omega \setminus \{y\}$.

Solution: True.

(d) [3 points] For an unconstrained optimization problem, if a point x satisfies the second order necessary condition, then it cannot be a saddle point.

Solution: False.

(e) [3 points] Consider $\min_{x\in\mathbb{R}^n} f(x)$, where $f(x) = x^\top Qx$ with $Q \in \mathbb{R}^{n\times n}$ being a positive definite matrix, then Newton's method with arbitrary initial point x^0 can converge to the optimal solution in one iteration.

Solution: True.

Question 2 [15 points]: KKT Conditions

Consider the following constrained optimization problem

$$\begin{array}{ll} \text{minimize} & 4x_1^2 + x_2^2 - x_1 - 2x_2 \\ \text{subject to} & x_1^2 \le 1. \end{array}$$

(a) [7 points] Derive the KKT conditions for this problem.

(b) [8 points] Find an optimal solution to this optimization problem.

Solution:

(a). We introduce dual variable $\lambda \ge 0$. Therefore the KKT conditions are listed as follows:

$$\begin{aligned} 8x_1 + 2\lambda x_1 - 1 &= 0 && \text{Main condition} && (1) \\ 2x_2 - 2 &= 0 && && (2) \\ \lambda &\ge 0 && \text{Dual feasibility} && (3) \\ \lambda(x_1^2 - 1) &= 0 && \text{Complementarity condition} && (4) \\ x_1^2 &\le 1 && \text{Primal feasibility} && (5) \end{aligned}$$

(b). Since the problom is a convex one, then KKT points must be globally optimal. By the main condition we have $x_2 = 1$. By complementarity condition, we have the following cases.

Case 1. $\lambda = 0$. Then from the main condition $x_1 = \frac{1}{8}$.

Case 2. $\lambda \neq 0$. Then $x_1 = 1$ or -1. When $x_1 = 1$, $\lambda = -\frac{7}{2}$. When $x_1 = -1$, $\lambda = -\frac{9}{2}$. In both of the cases, the dual feasibility is violated.

Therefore $x = (\frac{1}{8}, 1)$ is an optimal solution.

Question 3 [18 points]: Convexity

(a) [6 points] A mapping $g : \mathbb{R}^n \to \mathbb{R}^n$ is called monotone if

$$\langle g(x) - g(y), x - y \rangle \geq 0, \quad \forall x, y \in \mathbb{R}^n$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function. Show that its gradient ∇f is monotone.
Hint. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, it is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y \in \mathbb{R}^n$$

(b) [6 points] Let $x, y \in \mathbb{R}$. Given y is fixed, is the function $h(x) = (xy - 1)^2 + \frac{1}{2}(x - y)^2$ is a convex function? Given x is fixed, is $g(y) = (xy - 1)^2 + \frac{1}{2}(x - y)^2$ a convex function? Explain your answer.

(c) [6 points] Let $x, y \in \mathbb{R}$. Is $f(x, y) = (xy - 1)^2 + \frac{1}{2}(x - y)^2$ a convex function in (x, y)? Explain your answer.

Solution:

(a) From the convexity of f, we have that

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \text{ and } f(x) \geq f(y) + \nabla f(y)^\top (x - y).$$

Summing up these two equations,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

(b). Yes. This can be easily verified by checking PSDness of the Hessian matrices.

$$h''(x) = 2y^2 + 1 > 0$$

and

$$g''(y) = 2x^2 + 1 > 0.$$

(c) No. The Hessian matrix, which is given by

$$\begin{bmatrix} 2y^2 + 1 & 4xy - 3 \\ 4xy - 3 & 2x^2 + 1 \end{bmatrix}$$

is not PSD for some $(x, y) \in \mathbb{R}^2$. For instance at $(x, y) = (0, 0)$ the Hessian is $H = \begin{bmatrix} 1 & -3 \\ -3 & 1 \end{bmatrix}$. We have that $\mathrm{tr}(H) = 2 > 0$ and $\det(H) = -8 < 0$. Hence, H must be indefinite.

Question 4 [20 points]: Branch-and-Bound Method

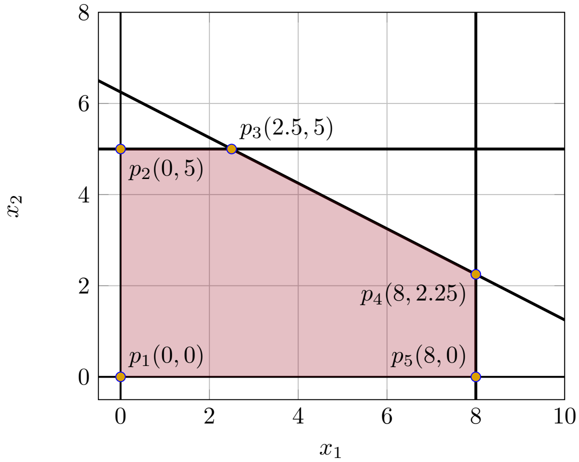
Use the branch-and-bound method to solve the following integer program.

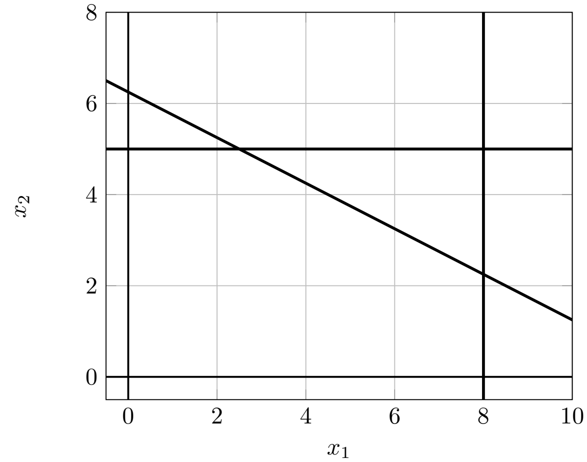
$$
\begin{array}{lll}
\text{maximize} & 12x_1 + 20x_2 & \\
\text{subject to} & 2x_1 + 4x_2 & \leq 25 \\
& x_1 & \leq 8 \\
& x_2 & \leq 5 \\
& x_1, x_2 & \geq 0 \\
& x, y \in \mathbb{Z}. &
\end{array}
$$

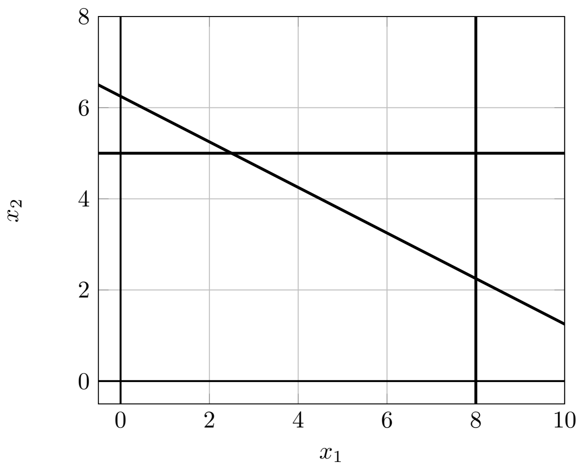
Please specify the branch-and-bound tree and what you did at each node.

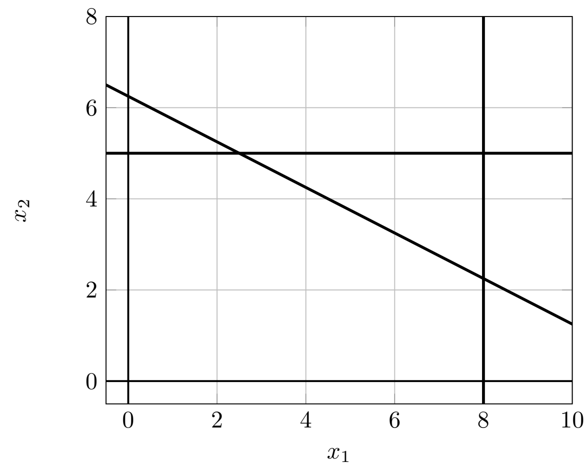
Hint: The following is the feasible region formed by the linear constraints of the original integer program, i.e.

$$\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : 2x_1 + 4x_2 \leq 25, x_1 \leq 8, x_2 \leq 5, x_1, x_2 \geq 0\}.$$









Solution:

We first draw the feasible set

$$\Omega := \{(x, y) \in \mathbb{R}^2 : 2x_1 + 4x_2 \leq 25, x_1 \leq 8, x_2 \leq 5, x_1 \geq 0, x_2 \geq 0\}$$

of the first LP relaxation.

The five extreme points of the polyhedron Ω are given by $p_1 = (0,0)^\top$, $p_2 = (0,5)^\top$, $p_3 = (2.5,5)^\top$, $p_4 = (8,2.25)^\top$, $p_5 = (8,0)^\top$. It is easy to see that the optimal solution of the relaxed LP is attained at p_4 with optimal value 141. This also means that the optimal function value of the integer program needs to be less or equal than 141.

We branch on $x_2 = 2.25$. We consider the two branches:

- (S1): $x_2 \leq 2$.
- (S2): $x_2 \geq 3$.

For (S1), the solution of the LP relaxation is given by $(8,2)^\top$ with objective value 136. This is an integer solution and we obtain the lower bound 136.

For (S2), using the sketch, the optimal solution satisfy $x_2 = 3$ and $2x_1 + 4x_2 = 25$. This gives $(6.5,3)^\top$ with optimal value 138.

We need to further branch on x_1. We consider the two branches:

- (S3): $x_1 \leq 6$.
- (S4): $x_1 \geq 7$.

We immediately see that (S4) is infeasible. For (S3), the extreme point with $x_1 = 6$ and $2x_1 + 4x_2 = 25$ needs to be the solution. This yields $(6,3.25)^\top$ and the corresponding function values is 139.

We continue branching on $x_2 = 3.25$:

- (S5): $x_2 \leq 3$
- (S6): $x_2 \geq 4$.

The optimal solution of the LP relaxation of (S5) is now $(6,3)^\top$ with optimal value 132, which is smaller than the lower bound 136. So we stop this branch.

The optimal solution of the LP relaxation of (S6) is given by $x_1 = 4.5$ and $x_2 = 4$ with objective value 134, which is smaller than lower bound 136. So we stop this branch.

Therefore, the optimal solution for this problem is (8,2), with objective value 136.

Question 5 [21 points]: Algorithm

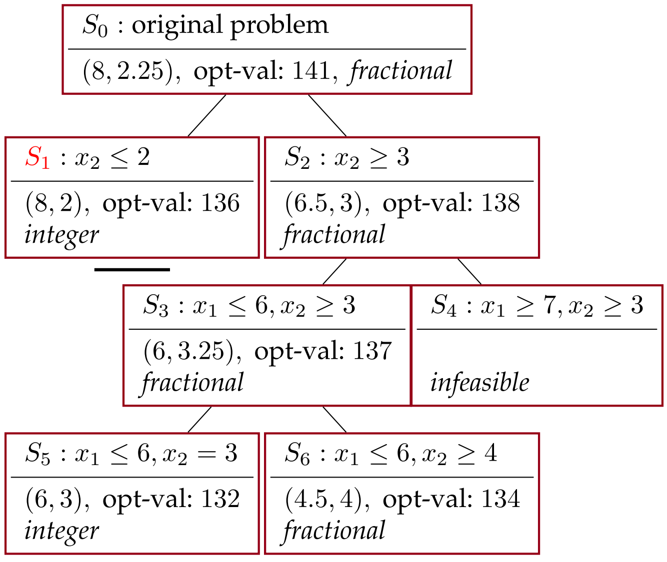
Consider the following unconstrained optimization problem:

$$\underset{x \in \mathbb{R}}{\text{minimize}} \ f(x),$$

where

$$f(x) = \begin{cases} \dfrac{3(1-x)^2}{4} - 2(1-x) & \text{if} \quad x > 1 \\ \dfrac{3(1+x)^2}{4} - 2(1+x) & \text{if} \quad x < -1 \\ x^2 - 1 & \text{if} \ -1 \leq x \leq 1 \end{cases}$$

Note that f is convex and continuously differentiable for all $x \in \mathbb{R}$.



(a) [3 points] Compute the gradient (derivative) of f.

Solution:

$$\nabla f(x) = \begin{cases} \frac{3x}{2} + \frac{1}{2} & \text{if} \quad x > 1 \\ \frac{3x}{2} - \frac{1}{2} & \text{if} \quad x < -1 \\ 2x & \text{if} \quad -1 \leq x \leq 1 \end{cases}$$

(b) [10 points] To apply gradient descent with line search for minimizing f, perhaps the simplest line search rule for choosing step size α^k is defined by the following condition

$$f(x^k + \alpha^k d^k) \leq f(x^k). \tag{LS}$$

The largest element in $\{1, \sigma, \sigma^2, \ldots\}$ satisfying the condition (LS) will be accepted as α^k, where $\sigma \in (0, 1)$.

Suppose we start at any x^0 satisfying $|x^0| > 1$. Show that the sequence of iterates $\{x^k\}$ generated by gradient descent with the line search (LS) will be bounded away from the optimal solutions. That is, it will never converge to an optimal solution, and hence this simple line search rule fails to minimize f.

Hint: A useful property of the function f: for any two scalers $x, y \in \mathbb{R}$, we have $f(x) < f(y)$ if and only if $|x| < |y|$.

Solution: Let $\nabla f(x) = 0$, we get $x^* = 0$. Since f is convex, thus x^* is the unique optimal solution. Furthermore, we can observe that for any two scalers $x, y \in \mathbb{R}$, we have

$$f(x) < f(y) \quad \text{if and only if} \quad |x| < |y|.$$

For all $x^0 > 1$, we have

$$x^0 - \nabla f(x^0) = -(1 + \frac{x^0 - 1}{2}).$$

It is easy to see that $x^0 - \nabla f(x^0) < -1$ and $|x^0 - \nabla f(x^0)| < |x^0|$. Thus, we have $f(x^0 - \nabla f(x^0)) < f(x^0)$.

For all $x^0 < -1$, we have

$$x^0 - \nabla f(x^0) = x^0 - \frac{3x^0}{2} + \frac{1}{2} = (1 - \frac{x^0+1}{2}).$$

It is easy to see that $x^0 - \nabla f(x^0) > 1$ and $|x^0 - \nabla f(x^0)| < |x^0|$. Thus, we have $f(x^0 - \nabla f(x^0)) < f(x^0)$.
In conclusion, the step size $\alpha^0 = 1$ will be accepted by the rule (LS) and the next point

$$x^1 = x^0 - \nabla f(x^0)$$

satisfies $|x^1| > 1$. By repeating such an argument infinitely, we will see that $|x^k| > 1$ for all $k \geq 0$, and hence never converge to the optimal solution $x^* = 0$.

(c) [8 points] Now, let us study the behavior of Armijo line search for minimizing f.
We choose the initial point

$$x^0 = \frac{3}{2}$$

and apply Armijo line search to select the step size α_0 with Armijo parameters set as follows:

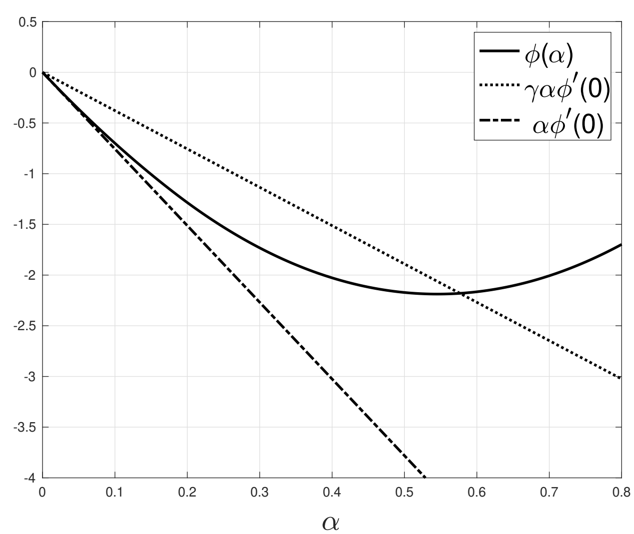
$$\gamma = \frac{1}{2}, \quad \sigma = \frac{1}{2}.$$

Now let us define

$$\phi(\alpha) := f(x^0 - \alpha\nabla f(x^0)) - f(x^0).$$

- Determine the step size α_0 using the Armijo line search and use such a α_0 to compute the next iterate x^1. Compared to the line search (LS) in part (b), what is your observation?
- Based on the value of x^1, will the sequence $\{x^k\}$ generated by gradient descent with Armijo line search converge to a minimizer of f? (You do not need to prove the convergence. You can just verify the conditions of the local convergence theorem provided in the slides.)

Hint: The following figure can be helpful for applying Armijo line search:



Solution:

- Armijo condition is satisfied whenever $\phi(\alpha) \le \gamma\alpha\phi'(0)$. Since, we choose $\sigma = \frac{1}{2}$, the backtracking procedure will terminate when $\alpha = \frac{1}{2}$.

  Since
  $$\nabla f(x^0) = \frac{11}{4}$$
  and the gradient descent iterates as:
  $$x^{k+1} = x^k - \alpha_k \nabla f(x^k).$$
  Thus, the next iterate will be
  $$x^1 = x^0 - \alpha_0 \nabla f(x^0) = \frac{1}{8}.$$
  Since $x^0 = \frac{3}{2} > 1$, if we use the simple line search rule (LS) in part (b), we will never have $|x^k| < 1$ for all $k \ge 0$. But if we use the Armijo line search, we have $|x^1| = \frac{1}{8} < 1$. Hence, Armijo line search overcomes this critical flaw.

- Recall that $x^1 = \frac{1}{8}$. Since f is continuously differentiable, Armijo line search is always well defined. Thus, we will have $f(x^k) < 0$ when $k \ge 1$ since Armijo line search will always decrease the function value and $f(x^1) = -\frac{63}{64} < 0$. Thus, we have $x^k \in [-1, 1]$ for all $k \ge 1$, otherwise $f(x^k) > 0$. In this region, we have
  $$\nabla f(x) = 2x,$$
  which is Lipschitz continuous with parameter $L = 2$. Furthermore, we have
  $$\nabla^2 f(x) = 2 > 0.$$
  We have verified all the conditions for local convergence theorem. Thus, $\{x^k\}$ will converge (linearly) to the optimal solution $x^* = 0$.

Question 6 [11 points]: Descent with Constant Step Size

Consider the following optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x),$$

where f is continuously differentiable and the gradient ∇f is L-Lipschitz continuous ($L > 0$), i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

In this question, we will show that gradient descent for minimizing f with a small constant step size will have descent property, i.e., line search is not required in this case.

(a) [6 points] Show that

$$|f(y) - f(x) - \nabla f(x)^\top (y - x)| \le \frac{L}{2}\|y - x\|^2 \quad x, y \in \mathbb{R}^n.$$

Hint: Consider the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(t) = f(x + t(y - x))$, where $x, y \in \mathbb{R}^n$ are constants. By Chain rule, we have $g'(t) = \nabla f(x + t(y - x))^\top (y - x)$. Then, you can start with the following Fundamental Theorem of Calculus

$$f(y) - f(x) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 \nabla f(x + t(y-x))^\top (y-x)dt.$$

Solution: We have

$$f(y) - f(x) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 \nabla f(x + t(y-x))^\top (y-x)dt.$$

Note that

$$\int_0^1 \nabla f(x+t(y-x))^\top (y-x)dt = \int_0^1 [\nabla f(x+t(y-x)) - \nabla f(x)]^\top (y-x)dt + \nabla f(x)^\top (y-x)$$

Coming the above two equations yields

$$\begin{aligned} |f(y) - f(x) - \nabla f(x)^\top (y-x)| &= \int_0^1 [\nabla f(x+t(y-x)) - \nabla f(x)]^\top (y-x)dt \\ &\le \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\| \|y-x\| dt \\ &\le \int_0^1 tL\|y-x\|^2 dt \\ &= L\|y-x\|^2 \int_0^1 tdt \\ &= \frac{L}{2}\|y-x\|^2, \end{aligned}$$

where the first and second inequalities are form Cauchy-Schwarz inequality and Lipschitz continuous gradient, respectively.

(b) [5 points] Based on part (a), suppose the current iterate x^k is not a stationary point, show that the gradient descent method $x^{k+1} = x^k - \alpha \nabla f(x^k)$ with a constant step size satisfying $\alpha < \frac{2}{L}$ can decrease the function value.

Solution: Set $y = x^{k+1}$ and $x = x^k$, from part (a), we have

$$f(x^{k+1}) \le f(x^k) + \nabla f(x^k)^\top (x^{k+1} - x^k) + \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

By realizing that $x^{k+1} - x^k = -\alpha \nabla f(x^k)$, we have

$$\begin{aligned} f(x^{k+1}) &\le f(x^k) - \alpha\|\nabla f(x^k)\|^2 + \frac{L\alpha^2}{2}\|\nabla f(x^k)\|^2 \\ &= f(x^k) - \alpha\left(1 - \frac{L\alpha}{2}\right)\|\nabla f(x^k)\|^2. \end{aligned}$$

Note that $\|\nabla f(x^k)\| > 0$ is x^k is not a stationary point. Once $\alpha < \frac{2}{L}$, we get descent of the function value.