

# MA3236 - Non-Linear Programming Suggested Solutions

(Semester 1, AY2022/2023)

Written by: Chow Yong Lam  
Audited by: Clarence Chew Xuan Da

## Question 1

(a)  $f(\mathbf{x}) = \frac{1}{4}x_1^4 + \frac{1}{4}x_2^4 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$ .

$$\nabla f(\mathbf{x}) = \begin{pmatrix} x_1^3 + x_1 \\ x_2^3 + x_2 \end{pmatrix}.$$

$$H_f(\mathbf{x}) = \begin{pmatrix} 3x_1^2 + 1 & 0 \\ 0 & 3x_2^2 + 1 \end{pmatrix}.$$

Since for all  $x_1, x_2 \in \mathbb{R}$ ,  $3x_1^2 + 1 \geq 1 > 0$ ,  $(3x_1^2 + 1)(3x_2^2 + 1) \geq 1(1) > 0$ , by the check of principal minors, the Hessian matrix  $H_f(\mathbf{x})$  is positive definite, and the function  $f$  is strictly convex.

(b)  $\mathbf{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - H_f(\mathbf{x}^{(0)})^{-1} \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1^3 + 1 \\ 1^3 + 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - H_f(\mathbf{x}^{(1)})^{-1} \nabla f(\mathbf{x}^{(1)}) = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \begin{pmatrix} \frac{7}{4} & 0 \\ 0 & \frac{7}{4} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2}^3 + \frac{1}{2} \\ \frac{1}{2}^3 + \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \begin{pmatrix} \frac{4}{7} & 0 \\ 0 & \frac{4}{7} \end{pmatrix} \begin{pmatrix} \frac{5}{8} \\ \frac{5}{8} \end{pmatrix} = \begin{pmatrix} \frac{1}{7} \\ \frac{1}{7} \end{pmatrix}. \text{ Note that}$$

$f(\mathbf{x}) \geq 0$  with equality achieved if and only if  $x_1 = x_2 = 0$ . Thus,  $\mathbf{x}^* = \mathbf{0}$ .

The Euclidean 2-norm of the error is:

$$\|\mathbf{x}^{(2)} - \mathbf{x}^*\| = \left\| \begin{pmatrix} \frac{1}{7} \\ \frac{1}{7} \end{pmatrix} \right\| = \left( 2 \left( \frac{1}{7} \right)^2 \right)^{\frac{1}{2}} = \frac{\sqrt{2}}{7}$$

(c) Given a function  $g$  and a point  $\mathbf{x} \in \mathbb{R}^n$ , a vector  $\mathbf{p}$  is a descent direction if  $\nabla g(\mathbf{x})^T \mathbf{p} < 0$ .

Since  $g$  is a strictly convex function, its Hessian matrix  $H_g(\mathbf{x})$  is a positive definite matrix at all  $\mathbf{x} \in \mathbb{R}^n$ , and so is the inverse  $H_g(\mathbf{x})^{-1}$ .

The Newton's direction of  $g$  at  $\mathbf{x}$  is  $-H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x})$ . If  $\mathbf{x}$  is a stationary point, then it is the global minimiser and the Newton's direction is just the zero vector. If  $\mathbf{x}$  is not a stationary point, then it is not the global minimiser and the gradient vector is non-zero. Thus,

$$\begin{aligned} \nabla g(\mathbf{x})^T (-H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x})) &= -\nabla g(\mathbf{x})^T H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}) \\ &< 0 \quad (\text{since } \nabla g(\mathbf{x})^T H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}) > 0) \end{aligned}$$

Therefore, the Newton's direction of  $g$  is a descent direction (when the point is not a global minimiser or stationary point).

**Note:** Alternatively, recall that  $H_g(\mathbf{x})$  is symmetric and so is  $H_g(\mathbf{x})^{-1}$ . In other words,  $H_g(\mathbf{x})^{-1} = (H_g(\mathbf{x})^{-1})^T$ . Since  $H_g(\mathbf{x})$  is positive definite,

$$\begin{aligned} \nabla g(\mathbf{x})^T H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}) &= \nabla g(\mathbf{x})^T (H_g(\mathbf{x})^{-1})^T H_g(\mathbf{x}) H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}) \\ &= (H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}))^T H_g(\mathbf{x}) H_g(\mathbf{x})^{-1} \nabla g(\mathbf{x}) \\ &> 0 \end{aligned}$$

## Question 2

$$\begin{aligned} \min \quad & f(\mathbf{x}) := -x_1 - x_2 \\ \text{s.t.} \quad & h_1(\mathbf{x}) := x_1^3 + 2x_1^2 + x_1 + x_2 - 1 \leq 0 \text{ and} \\ & h_2(\mathbf{x}) := -5x_1 - x_2 - 7 \leq 0 \end{aligned}$$

(a)

$$\nabla f(\mathbf{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}; \quad \nabla h_1(\mathbf{x}) = \begin{pmatrix} 3x_1^2 + 4x_1 + 1 \\ 1 \end{pmatrix}; \quad \nabla h_2(\mathbf{x}) = \begin{pmatrix} -5 \\ -1 \end{pmatrix}$$

Since both gradient vectors are non-zero, any feasible point with only one active inequality constraint is regular. Any interior point in the feasible region (no active constraints) is also regular.

If both inequality constraints are active, then:

$$\begin{aligned} h_2(\mathbf{x}) = 0 &\implies x_2 = -5x_1 - 7 \\ h_1(\mathbf{x}) = 0 &\implies x_1^3 + 2x_1^2 + x_1 - 5x_1 - 7 - 1 = 0 \\ &\implies (x_1 + 2)^2(x_1 - 2) = 0 \\ &\implies x_1 = -2 \text{ or } x_1 = 2 \end{aligned}$$

If  $x_1 = -2$ , then  $x_2 = 3$ ,  $\nabla h_1(\mathbf{x}) = \begin{pmatrix} 5 \\ 1 \end{pmatrix} = -\nabla h_2(\mathbf{x})$  and the two gradients are linearly dependent.

If  $x_1 = 2$ , then  $x_2 = -17$ ,  $\nabla h_1(\mathbf{x}) = \begin{pmatrix} 21 \\ 1 \end{pmatrix}$ , which is not a scalar multiple of  $\nabla h_2(\mathbf{x})$ . The two gradients are linearly independent.

Thus, all feasible points, except  $\mathbf{x} = [-2; 3]$ , are regular.

(b) Case 1: Any interior point is not a KKT point since  $\nabla f(\mathbf{x})$  is non-zero.

Case 2: Only  $h_1(\mathbf{x}) = 0$  is active, i.e.  $x_1^3 + 2x_1^2 + x_1 + x_2 - 1 = 0$ . If the point is a KKT point, then there exists  $\mu_1 \geq 0$  such that:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \mu_1 \begin{pmatrix} 3x_1^2 + 4x_1 + 1 \\ 1 \end{pmatrix} = 0$$

Then,  $\mu_1 = 1$  from the second equality. From the first equality,

$$\begin{aligned} 3x_1^2 + 4x_1 + 1 - 1 &= 0 \implies x_1(3x_1 + 4) = 0 \\ &\implies x_1 = 0 \quad \text{or} \quad x_1 = -\frac{4}{3} \end{aligned}$$

When  $x_1 = 0$ ,  $x_2 = 1$ . When  $x_1 = -\frac{4}{3}$ ,  $x_2 = 1 - x_1 - 2x_1^2 - x_1^3 = \frac{31}{27}$ . Therefore, there are two potential KKT points,  $[0; 1]$  and  $[-\frac{4}{3}; \frac{31}{27}]$ . Indeed,  $h_2([0; 1]) = -8 < 0$ ,  $h_2(-\frac{4}{3}; \frac{31}{27}) = -\frac{40}{27} < 0$ . Thus, both points are feasible and they are KKT points.

Case 3: Only  $h_2(\mathbf{x}) = 0$  is active. If the point is a KKT point, then there exists  $\mu_2 \geq 0$  such that:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} -5 \\ -1 \end{pmatrix} = 0$$

From the second equality,  $\mu_2 + 1 = 0 \implies \mu_2 = -1 < 0$ , a contradiction. Thus, there are no KKT points in this case.

Case 4: Both constraints are active. The only regular point in this case is  $[2; -17]$ . If it is a KKT point, then there exist  $\mu_1, \mu_2 \geq 0$  such that:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \mu_1 \begin{pmatrix} 21 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} -5 \\ -1 \end{pmatrix} = 0$$

By solving the equalities,  $\mu_1 = -\frac{1}{4}$ ,  $\mu_2 = -\frac{5}{4}$ . Both parameters are negative, so there are no KKT points in this case.

Therefore, there are two KKT points,  $[0; 1]$  and  $[-\frac{4}{3}; \frac{31}{27}]$ .

- (c) Consider the point  $\mathbf{x} = [t; -5t - 7], t \in \mathbb{R}$ . At this point, the condition  $h_2(\mathbf{x}) = 0$  holds. Then as  $t \rightarrow -\infty, h_1(\mathbf{x}) = (t + 2)^2(t - 2) \rightarrow -\infty$ . In particular,  $h_1(\mathbf{x}) < 0$  as long as  $t < -2$ . Using this point,  $f(\mathbf{x}) = -t + 5t + 7 = 4t + 7 \rightarrow -\infty$  as  $t \rightarrow -\infty$  from  $-2$ . Thus, the objective function value is unbounded in the feasible set and this problem does not have an optimal solution.

### Question 3

In this problem,  $a, b \in \mathbb{R}, a \neq 0$ .

$$(P) \quad \begin{aligned} \min \quad & f(x_1) \\ \text{s.t.} \quad & ax_2 - b = 0, \text{ and} \\ & x_3^2 - 1 \leq 0. \end{aligned}$$

- (a) Since  $f$  is a differentiable convex function, any stationary point of  $f$  is a global minimiser. In particular,  $x_1^*$  is a stationary point and thus a global minimiser.  
In this problem, any feasible point must fulfill  $x_2 = \frac{b}{a}, x_3 \in [-1, 1]$ . Thus, the optimal solutions are in the form  $\mathbf{x} = [x_1; \frac{b}{a}; x_3]$  where  $x_1$  is a stationary point of  $f$  and  $x_3 \in [-1, 1]$ . The optimal objective function value of (P) is  $f(x_1^*)$ .
- (b) For  $\lambda \in \mathbb{R}, \mu \geq 0$ , the Lagrangian function is given as:

$$L(\mathbf{x}, \lambda, \mu) = f(x_1) + \lambda(ax_2 - b) + \mu(x_3^2 - 1)$$

and the Lagrangian dual function is:

$$\begin{aligned} \theta(\lambda, \mu) &= \inf_{\mathbf{x} \in \mathbb{R}^3} \{f(x_1) + \lambda(ax_2 - b) + \mu(x_3^2 - 1)\} \\ &= \inf_{x_1 \in \mathbb{R}} \{f(x_1)\} + \inf_{x_2 \in \mathbb{R}} \{\lambda(ax_2 - b)\} + \inf_{x_3 \in \mathbb{R}} \{\mu(x_3^2 - 1)\} \end{aligned}$$

Note that:

$$\inf_{x_2 \in \mathbb{R}} \{\lambda(ax_2 - b)\} = \begin{cases} -\infty & \text{if } \lambda \neq 0 \\ 0 & \text{if } \lambda = 0 \end{cases}$$

and:

$$\inf_{x_3 \in \mathbb{R}} \{\mu(x_3^2 - 1)\} = -\mu$$

Thus,

$$\theta(\lambda, \mu) = \begin{cases} -\infty & \text{if } \lambda \neq 0 \\ f(x_1^*) - \mu & \text{if } \lambda = 0 \end{cases}$$

- (c) The Lagrangian dual problem is:

$$(D) \quad \begin{aligned} \max_{\lambda, \mu \in \mathbb{R}} \quad & f(x_1^*) - \mu \\ \text{s.t.} \quad & \lambda = 0, \text{ and} \\ & \mu \geq 0. \end{aligned}$$

The dual optimal solution is  $(\lambda, \mu) = (0, 0)$  and the dual optimal objective function value is  $f(x_1^*)$ .

- (d) There is zero duality gap since the primal and dual optimal objective function values are equal.
- (e) First, the objective function  $f$  is convex since the Hessian matrix as shown below is positive semi definite.

$$\forall \mathbf{x} \in \mathbb{R}^3, H(\mathbf{x}) = \begin{pmatrix} f''(x_1) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ where for all } x_1, f''(x_1) \geq 0.$$

$g(\mathbf{x}) = ax_2 - b$  is affine,  $h(\mathbf{x}) = x_3^2 - 1$  is convex so the feasible region is a convex set. Finally, using the point  $[x_1^*; \frac{b}{a}; 0]$ ,  $0 - 1 < 0, a(\frac{b}{a}) - b = 0$  so the Slater's condition holds. Thus, the conditions for the strong duality theorem hold.

## Question 4

In this problem,  $\mathbf{c} \in \mathbb{R}^n$ ,  $\mathbf{c} \geq 0$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ .

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{R}^n} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} \leq \mathbf{b}, \text{ and} \\ & \mathbf{x} \in X := \{0, 1\}^n \end{aligned}$$

(a) The Lagrangian dual function is, for  $\boldsymbol{\mu} \in \mathbb{R}^m$ ,  $\boldsymbol{\mu} \geq 0$ :

$$\begin{aligned} \theta(\boldsymbol{\mu}) &= \inf_{\mathbf{x} \in X} \mathbf{c}^T \mathbf{x} + \boldsymbol{\mu}^T (A\mathbf{x} - \mathbf{b}) \\ &= \inf_{\mathbf{x} \in X} (A^T \boldsymbol{\mu} + \mathbf{c})^T \mathbf{x} - \boldsymbol{\mu}^T \mathbf{b} \end{aligned}$$

Let  $\mathbf{p} = A^T \boldsymbol{\mu} + \mathbf{c}$ , and denote  $p_i$  as the  $i$ -th entry in  $\mathbf{p}$ . Then,

$$\theta(\boldsymbol{\mu}) = \sum_{i: p_i \leq 0} p_i - \boldsymbol{\mu}^T \mathbf{b}$$

The first sum can be obtained by choosing  $x_i = 1$  if and only if  $p_i \leq 0$ .

(b)  $\theta(\mathbf{0}) = \inf_{\mathbf{x} \in X} \mathbf{c}^T \mathbf{x}$ . Since  $\mathbf{c} \geq 0$ , then  $\theta(\mathbf{0}) \geq 0$ .

Case 1: If  $\mathbf{c}$  has no zero entries. Then the unique minimiser of  $\mathbf{c}^T \mathbf{x}$  in  $X$  is  $\mathbf{x} = \mathbf{0}$ , in which the minimum value is zero. Then  $\partial\theta(\mathbf{0}) = \text{conv}\{A \cdot \mathbf{0} - \mathbf{b}\} = \{-\mathbf{b}\}$ .

Case 2: Suppose there are entries in  $\mathbf{c}$  with value zero. Denote  $c_i$  as the  $i$ -th entry of  $\mathbf{c}$ . Then,  $\mathbf{c}^T \mathbf{x} = 0$  for all  $\mathbf{x}$  in the set below:

$$S = \{\mathbf{x} \in X : x_i \in \{0, 1\} \text{ for } i \text{ such that } c_i = 0, x_i = 0 \text{ for } i \text{ such that } c_i \neq 0\}$$

The set  $S$  contains all minimisers of  $\mathbf{c}^T \mathbf{x}$ , and  $\partial\theta(\mathbf{0}) = \text{conv}\{A\mathbf{x} - \mathbf{b} : \mathbf{x} \in S\}$ . Denote  $\mathbf{a}_i$  as the  $i$ -th column of  $A$ . The subdifferential can also be written as:

$$\partial\theta(\mathbf{0}) = \text{conv} \left\{ \sum_{i: c_i = 0} y_i \mathbf{a}_i - \mathbf{b} : y_i \in \{0, 1\} \text{ for } i \in \{1, 2, \dots, n\} \text{ in which } c_i = 0 \right\}$$

## Question 5

For (a) and (b):

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^3} \quad & x_1^2 + x_2^2 + x_3^2 - 2x_3 \\ \text{s.t.} \quad & 2x_1 + x_2 = 1 \end{aligned}$$

(a) The quadratic penalty function is:

$$Q(\mathbf{x}, \mu) = x_1^2 + x_2^2 + x_3^2 - 2x_3 + \frac{1}{2\mu}(2x_1 + x_2 - 1)^2$$

The function is convex, so the minimiser is the stationary point.

$$\nabla Q(\mathbf{x}, \mu) = \begin{pmatrix} 2x_1 + \frac{2}{\mu}(2x_1 + x_2 - 1) \\ 2x_2 + \frac{1}{\mu}(2x_1 + x_2 - 1) \\ 2x_3 - 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The minimizer is thus:

$$\mathbf{x} = \begin{pmatrix} \frac{2}{2\mu+5} \\ \frac{1}{2\mu+5} \\ 1 \end{pmatrix}$$

(b) The augmented Lagrangian function is:

$$L_A(\mathbf{x}, \lambda, \sigma) = x_1^2 + x_2^2 + x_3^2 - 2x_3 + \lambda(2x_1 + x_2 - 1) + \frac{1}{2\sigma}(2x_1 + x_2 - 1)^2$$

The function is convex, so the minimiser is the stationary point.

$$\nabla L_A(\mathbf{x}, \lambda, \sigma) = \begin{pmatrix} 2x_1 + 2\lambda + \frac{2}{\sigma}(2x_1 + x_2 - 1) \\ 2x_2 + \lambda + \frac{1}{\sigma}(2x_1 + x_2 - 1) \\ 2x_3 - 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The minimizer is thus:

$$\mathbf{x} = \begin{pmatrix} \frac{2(1-\lambda\sigma)}{2\sigma+5} \\ \frac{1-\lambda\sigma}{2\sigma+5} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix} \text{ when } \lambda = -\frac{2}{5}$$

For (c) and (d),  $Q \in \mathbb{R}^{n \times n}$  is symmetric positive definite,  $\mathbf{c} \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ .

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

(c) The quadratic penalty function is:

$$Q(\mathbf{x}, \mu) = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x} + \frac{1}{2\mu} \|A\mathbf{x} - \mathbf{b}\|^2$$

The function is convex, so the minimiser is the stationary point.

$$\nabla Q(\mathbf{x}, \mu) = Q\mathbf{x} + \mathbf{c} + \frac{1}{\mu} A^T (A\mathbf{x} - \mathbf{b}) = \mathbf{0}$$

Thus we need to solve:

$$\left( \frac{1}{\mu} A^T A + Q \right) \mathbf{x} = \frac{1}{\mu} A^T \mathbf{b} - \mathbf{c}$$

(d) The augmented Lagrangian function is:

$$L_A(\mathbf{x}, \boldsymbol{\lambda}, \sigma) = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x} + \boldsymbol{\lambda}^T (A\mathbf{x} - \mathbf{b}) + \frac{1}{2\sigma} \|A\mathbf{x} - \mathbf{b}\|^2$$

The function is convex, so the minimiser is the stationary point.

$$\nabla L_A(\mathbf{x}, \boldsymbol{\lambda}, \sigma) = Q\mathbf{x} + \mathbf{c} + A^T \boldsymbol{\lambda} + \frac{1}{\sigma} A^T (A\mathbf{x} - \mathbf{b}) = \mathbf{0}$$

Thus we need to solve:

$$\left( \frac{1}{\sigma} A^T A + Q \right) \mathbf{x} = \frac{1}{\sigma} A^T \mathbf{b} - \mathbf{c} - A^T \boldsymbol{\lambda}$$