NATIONAL UNIVERSITY OF SINGAPORE
MATHEMATICS SOCIETY

PAST YEAR PAPER SOLUTIONS
solutions prepared by Joseph Andreas

MA3236 Nonlinear Programming
AY 2010/2011 Sem 1

Question 1

(a) We have
$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + 3x_2^3 - x_2 \\ 9x_1x_2^2 - x_1 \end{pmatrix}$$
so then we have
$$H_f(\mathbf{x}) = \begin{bmatrix} 2 & 9x_2^2 - 1 \\ 9x_2^2 - 1 & 18x_1x_2 \end{bmatrix}$$
as desired.

(b) Suppose $\mathbf{x}^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ be a stationary point of f on S. Since $\nabla f(\mathbf{x}^*) = \begin{pmatrix} 2x_1 + 3x_2^3 - x_2 \\ 9x_1x_2^2 - x_1 \end{pmatrix}$, it follows that
$$2x_1 + 3x_2^3 - x_2 = 0 \tag{1}$$
$$9x_1x_2^2 - x_1 = 0 \tag{2}$$
From (2) we have either $x_1 = 0$ or $x_2 = \frac{1}{3}$ (note that $x_2 > 0$). If $x_1 = 0$, we see that $3x_2^3 - x_2 = 0$. Since $x_2 > 0$, we have $x_2 = \sqrt{\frac{1}{3}}$. If $x_2 = \frac{1}{3}$, then substitute to (1), we have $x_1 = \frac{1}{9}$. Thus, all stationary points of f in S are given by $\left(0, \sqrt{\frac{1}{3}}\right)$ and $\left(\frac{1}{9}, \frac{1}{3}\right)$

(c)
- $\mathbf{x} = \left(0, \sqrt{\frac{1}{3}}\right)$. We have $H_f(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}$. Hence if a_1, a_2 are two eigenvalues of $H_f(\mathbf{x})$, we have $a_1a_2 =| H_f(\mathbf{x}) |= 2 \times 0 - 2 \times 2 = -4 < 0$. Thus, a_1 and a_2 are on different sign, so we have $H_f(\mathbf{x})$ is indefinite. Therefore, the point $\left(0, \sqrt{\frac{1}{3}}\right)$ is a saddle point.
- $\mathbf{x} = \left(\frac{1}{9}, \frac{1}{3}\right)$. We have $H_f(\mathbf{x}) = \begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{3} \end{bmatrix}$ which is clearly positive definite since both eigenvalues 2 and $\frac{2}{3}$ are positive. Therefore, the point $\left(\frac{1}{9}, \frac{1}{3}\right)$ is a local minimizer.

Question 2

Since $B_r(\mathbf{0})$ is a nonempty (note that $r > 0$) closed and bounded set and f is continuous over $B_r(\mathbf{0})$, Weierstrass theorem implies that there is a global minimizer $\mathbf{x}^*$ of f over $B_r(\mathbf{0})$. Since $\hat{x} \in B_r(\mathbf{0})$ it follows that $f(\mathbf{x}^*) \leq f(\hat{x}) < \beta$. Now, take any $\mathbf{x} \in D$. If $\mathbf{x} \notin B_r(\mathbf{0})$, we have $f(\mathbf{x}) \geq \beta > f(\mathbf{x}^*)$. If $\mathbf{x} \in B_r(\mathbf{0})$, then $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ since $\mathbf{x}^*$ is a global minimizer of f over $B_r(\mathbf{0})$. We conclude $\mathbf{x}^*$ is a global minimizer over f on entire D. QED.

Question 3

(a) We have

$$g(\mathbf{x}) \geq \frac{\rho}{2}||\mathbf{x}+\frac{\mathbf{b}}{\rho}||^2 + f(\mathbf{0}) - \frac{1}{2\rho}||\mathbf{b}||^2$$

$$\Leftrightarrow f(\mathbf{x}) + \frac{\rho}{2}\mathbf{x}^T\mathbf{x} \geq \frac{\rho}{2}\left(\mathbf{x}+\frac{\mathbf{b}}{\rho}\right)^T\left(\mathbf{x}+\frac{\mathbf{b}}{\rho}\right) + f(\mathbf{0}) - \frac{1}{2\rho}\mathbf{b}^T\mathbf{b}$$

$$\Leftrightarrow f(\mathbf{x}) + \frac{\rho}{2}\mathbf{x}^T\mathbf{x} \geq \frac{\rho}{2}\mathbf{x}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{b} + \frac{1}{2}\mathbf{b}^T\mathbf{x} + \frac{1}{2\rho}\mathbf{b}^T\mathbf{b} + f(\mathbf{0}) - \frac{1}{2\rho}\mathbf{b}^T\mathbf{b}$$

$$\Leftrightarrow f(\mathbf{x}) \geq \nabla\mathbf{f}(\mathbf{0})^T\mathbf{x} + f(\mathbf{0})$$

Hence, the inequality given in the problem is equivalent to the last inequality. However, the last inequality is true since f is convex, by tangent plane characterization.
Now since $\lim_{||\mathbf{x}||\to\infty}||\mathbf{x}+\frac{1}{\rho}\mathbf{b}||^2 = \infty$, it's easy to see that g is coercive. QED.

(b) Suppose $\mathbf{u}$ and $\mathbf{v}$ are two vectors. We consider any α from interval $(0,1)$. Then we have $\alpha q(\mathbf{u}) + (1-\alpha)q(\mathbf{v}) = \alpha h(A\mathbf{u}+\mathbf{b}) + (1-\alpha)h(A\mathbf{v}+\mathbf{b}) \geq h(\alpha(A\mathbf{u}+\mathbf{b}) + (1-\alpha)(A\mathbf{v}+\mathbf{b})) = h(A(\alpha\mathbf{u}+(1-\alpha)\mathbf{v})+\mathbf{b}) = q(\alpha\mathbf{u}+(1-\alpha)\mathbf{v})$. By definition, q is convex. Note that the inequality comes from the convexity of h. QED.
On the other hand, assume h is strictly convex. We shall show that q need not be strictly convex. Take an example if $nullity(A) > 0$. Clearly, such matrix A exists. Then $A\mathbf{x} = 0$ has infinitely non-zero solutions. Take two different non-zero solutions $\mathbf{u}$ and $\mathbf{v}$. Note that $A\mathbf{u} = A\mathbf{v} = 0$. Thus, we have $A(\frac{\mathbf{u}+\mathbf{v}}{2}) = 0$. Hence, $\frac{q(\mathbf{u})+q(\mathbf{v})}{2} = \frac{h(A\mathbf{u}+\mathbf{b})+h(A\mathbf{v}+\mathbf{b})}{2} = h(\mathbf{b}) = h\left(A\left(\frac{\mathbf{u}+\mathbf{v}}{2}\right)+\mathbf{b}\right) = q\left(\frac{\mathbf{u}+\mathbf{v}}{2}\right)$
Hence, q may not be strictly convex.

Question 4

(a) claim : $\lambda_{\min(\mathbf{Q})}||\mathbf{x}||^2 \leq \mathbf{x}^T\mathbf{Q}\mathbf{x} \leq \lambda_{\max(\mathbf{Q})}||\mathbf{x}||^2$
Proof of claim : (no need to write the proof of this claim in the exam, just use it) Since $\mathbf{Q}$ is symmetric, then there exists a $n\times n$ orthogonal matrix $\mathbf{P}$ such that $\mathbf{D} = \mathbf{P}^T\mathbf{Q}\mathbf{P}$ for some diagonal matrix $\mathbf{D}$. (note that, it also mean $\mathbf{Q} = \mathbf{P}\mathbf{D}\mathbf{P}^T$
Since $\mathbf{P}$ is orthogonal, it follows that $\mathbf{P}^T = \mathbf{P}^{-1}$. Hence, $\mathbf{P}$ is invertible and from there, we get the row vectors of $\mathbf{P}$ are the basis of our $\mathbb{R}^n$. Therefore, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are row vectors of $\mathbf{P}$, we have the set $\{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \mid c_1, c_2, \ldots, c_n \in \mathbb{R}\}$ is actually equal to $\mathbb{R}^n$. Therefore, we conclude,

$$\{\mathbf{P}\mathbf{u} \mid \mathbf{u} \in \mathbb{R}^n\} = \mathbb{R}^n$$

Notice that if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all eigenvalues of Q, it follows that

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

We pick an arbitrary vector in $\mathbb{R}^n$, i.e $\mathbf{x}$. Since $\{\mathbf{P}\mathbf{u} \mid \mathbf{u} \in \mathbb{R}^n\} = \mathbb{R}^n$, we get there is a vector $\mathbf{y}$ such that $\mathbf{P}\mathbf{y} = \mathbf{x}$. Hence, $\mathbf{x}^T\mathbf{Q}\mathbf{x} = (\mathbf{P}\mathbf{y})^T\mathbf{Q}\mathbf{P}\mathbf{y} = \mathbf{y}^T\mathbf{P}^T\mathbf{Q}\mathbf{P}\mathbf{y} = \mathbf{y}^T\mathbf{D}\mathbf{y}$. Furthermore, if $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix}^T$, we have $\mathbf{y}^T\mathbf{D}\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2$. Since $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all positive and $\lambda_{\min(\mathbf{Q})} \leq \lambda_i \leq \lambda_{\max(\mathbf{Q})}$ for each i, it follows that
$\lambda_{\min(\mathbf{Q})}\left(y_1^2 + y_2^2 + \ldots + y_n^2\right) \leq \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2 \leq \lambda_{\max(\mathbf{Q})}\left(y_1^2 + y_2^2 + \ldots + y_n^2\right)$, or equivalently, we get $\lambda_{\min(\mathbf{Q})}\mathbf{y}^T\mathbf{y} \leq \mathbf{y}^T\mathbf{D}\mathbf{y} \leq \lambda_{\max(\mathbf{Q})}\mathbf{y}^T\mathbf{y}$. We recall some facts first $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x}$ and $\mathbf{P}^T = \mathbf{P}^{-1}$ which leads $\mathbf{y} = \mathbf{P}^T\mathbf{x}$. Another fact : $\mathbf{P}\mathbf{P}^T = \mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$.
Finally,

$$\lambda_{\min(\mathbf{Q})}\mathbf{y}^T\mathbf{y} \leq \mathbf{y}^T\mathbf{D}\mathbf{y} \leq \lambda_{\max(\mathbf{Q})}\mathbf{y}^T\mathbf{y}$$

$$\Leftrightarrow \lambda_{\min(\mathbf{Q})}(\mathbf{P}^T\mathbf{x})^T\mathbf{P}^T\mathbf{x} \le (\mathbf{P}^T\mathbf{x})^T\mathbf{D}\mathbf{P}^T\mathbf{x} \le \lambda_{\max(\mathbf{Q})}(\mathbf{P}^T\mathbf{x})^T\mathbf{P}^T\mathbf{x}$$
$$\Leftrightarrow \lambda_{\min(\mathbf{Q})}\mathbf{x}^T(\mathbf{P}^T)^T\mathbf{P}^T\mathbf{x} \le \mathbf{x}^T(\mathbf{P}^T)^T\mathbf{D}\mathbf{P}^T\mathbf{x} \le \lambda_{\max(\mathbf{Q})}\mathbf{x}^T(\mathbf{P}^T)^T\mathbf{P}^T\mathbf{x}$$
$$\Leftrightarrow \lambda_{\min(\mathbf{Q})}\mathbf{x}^T\mathbf{P}\mathbf{P}^T\mathbf{x} \le \mathbf{x}^T\mathbf{P}\mathbf{D}\mathbf{P}^T\mathbf{x} \le \lambda_{\max(\mathbf{Q})}\mathbf{x}^T\mathbf{P}\mathbf{P}^T\mathbf{x}$$
$$\Leftrightarrow \lambda_{\min(\mathbf{Q})}\mathbf{x}^T\mathbf{x} \le \mathbf{x}^T\mathbf{Q}\mathbf{x} \le \lambda_{\max(\mathbf{Q})}\mathbf{x}^T\mathbf{x}$$

Claim is proved.

From our claim, we get $\frac{1}{\lambda_{\max(\mathbf{Q})}||\mathbf{x}||^2} \le \frac{1}{\mathbf{x}^T\mathbf{Q}\mathbf{x}} \le \frac{1}{\lambda_{\min(\mathbf{Q})}||\mathbf{x}||^2}$. Furthermore, we also get $\forall \mathbf{x} \in L_\alpha$, $\frac{1}{\lambda_{\max(\mathbf{Q})}||\mathbf{x}||^2} \le \frac{1}{\alpha} \le \frac{1}{\lambda_{\min(\mathbf{Q})}||\mathbf{x}||^2}$. Multiplying the inequality by a positive number $\alpha||x||^2$, we get $\frac{\alpha}{\lambda_{\max(\mathbf{Q})}} \le ||x||^2 \le \frac{\alpha}{\lambda_{\min(\mathbf{Q})}}$ as desired. QED.

(b) We have $\kappa(\mathbf{Q}) = \left(\frac{\lambda_{\max(\mathbf{Q})}}{\lambda_{\min(\mathbf{Q})}}\right)^2 = \left(\frac{2.8\text{cm}}{0.7\text{cm}}\right)^2 = 16$. (the value 2.8 cm is the major axis, and 0.7 cm is minor axis, one can also prove that $\frac{\lambda_{max}}{\lambda_{min}} = \frac{\text{major axis}}{\text{minor axis}}$). Hence, the rate of convergence is given by $\rho(\mathbf{Q}) = \left(\frac{16-1}{16+1}\right)^2 = \frac{225}{289}$. Since $\epsilon = 10^{-8}$, we get the number of iterations needed is given by $k = \left\lceil \frac{\log e}{\log \rho(\mathbf{Q})} \right\rceil + 1 = 75$

(c) The solution t^* is the solution of the following problem :

$$\min g(t) = f(\mathbf{x} + t\mathbf{p}) \text{ with } t \ge 0$$

It happens that $g'(t^*) = 0$. Note also $g'(t) = \langle \nabla f(\mathbf{y} + t\mathbf{p}), \mathbf{p}\rangle$. Furthermore, since f is quadratic, it's well known that $\nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} + \mathbf{c}$. It follows that :

$$\begin{aligned} 0 = g'(t^*) &= \langle \nabla f(\mathbf{y} + t^*\mathbf{p}), \mathbf{p}\rangle \\ &= \langle \mathbf{Q}(\mathbf{y} + t^*\mathbf{p}) + \mathbf{c}, \mathbf{p}\rangle \\ &= \langle \mathbf{Q}\mathbf{y} + \mathbf{c} + t^*\mathbf{Q}\mathbf{p}, \mathbf{p}\rangle \\ &= \langle \mathbf{Q}\mathbf{y} + \mathbf{c}, \mathbf{p}\rangle + t^*\langle \mathbf{Q}\mathbf{p}, \mathbf{p}\rangle \\ &= \langle \nabla f(\mathbf{y}), \mathbf{p}\rangle + t^*\langle \mathbf{Q}\mathbf{p}, \mathbf{p}\rangle \end{aligned}$$

From here, we get $t^* = -\frac{\langle \nabla f(\mathbf{y}), \mathbf{p}\rangle}{\langle \mathbf{Q}\mathbf{p}, \mathbf{p}\rangle} = -\frac{\nabla f(\mathbf{y})^T\mathbf{p}}{\mathbf{p}^T\mathbf{Q}\mathbf{p}}$.

Question 5

Before we begin, let us change the problem a little bit into our regular NLP. The problem is :

$$\begin{array}{rl} \min & f(\mathbf{x}) := x_1 + x_2 \\ \text{s.t} & h_1(\mathbf{x}) := -(x_1+1)^2 - x_2^2 \le -1 \\ & h_2(\mathbf{x}) := x_1^2 + 2x_2^2 \le 3 \end{array}$$

We shall solve this equivalent NLP problem for our remaining problem instead of the original one. Some facts :

(i) $\nabla f(\mathbf{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(ii) $\nabla h_1(\mathbf{x}) = \begin{pmatrix} -2x_1 - 2 \\ -2x_2 \end{pmatrix}$

(iii) $\nabla h_2(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$

(iv) $H_f(\mathbf{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

(v) $H_L(\mathbf{x}) = \mu_1 \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} + \mu_2 \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$

(a) Let $\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix}^T$ be an arbitrary feasible point.

Case 1 None of h_1, h_2 are active at $\mathbf{x}$. Clearly, $\mathbf{x}$ is regular in this case.

Case 2 h_1 is active, but h_2 is not. Then we have $(x_1+1)^2 + x_2^2 = 1$. Hence, $\mathbf{x} \neq \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ as otherwise, the previous equality would not be satisfied. However, looking back at (ii), we know that $\nabla h_1(\mathbf{x}) = 0$ iff $\mathbf{x} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, so then we have $\nabla h_1(\mathbf{x}) \neq 0$. Hence, the set $J(\mathbf{x}) = \{\nabla h_1(\mathbf{x})\}$ is linearly independent. Therefore, x is regular.

Case 3 h_2 is active, but h_1 is not. Then we have $x_1^2 + 2x_2^2 = 3$. Hence, $\mathbf{x} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ as otherwise, the previous equality would not be satisfied. However, looking back at (iii), we know that $\nabla h_2(\mathbf{x}) = 0$ iff $\mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, so then we have $\nabla h_2(\mathbf{x}) \neq 0$. Hence, the set $J(\mathbf{x}) = \{\nabla h_2(\mathbf{x})\}$ is linearly independent. Therefore, x is regular.

Case 4 Both h_1 and h_2 are active. Hence we have

$$(x_1+1)^2 + x_2^2 = 1 \tag{1}$$
$$x_1^2 + 2x_2^2 = 3 \tag{2}$$

Now, multiply (1) by 2 and subtract it by (2), we get $2(x_1+1)^2 - x_1^2 = -1$ which is equivalent with $x_1^2 + 4x_1 + 3 = 0 \Leftrightarrow x_1 = -1$ or $x_1 = -3$. If $x_1 = -1$, substitute back to (2), we get $x_2^2 = 1$, so $x_2 = \pm 1$. If $x_1 = -3$, substitute back to (2) we get $x_2^2 = -3$, a contradiction. Hence, if h_1, h_2 is active, then $x_1 = 1$ and $x_2 = \pm 1$.........................(3)

For $x_1 = 1$ and $x_2 = 1$, we get $\nabla h_1(\mathbf{x}) = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$ and $\nabla h_2(\mathbf{x}) = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$

For $x_1 = 1$ and $x_2 = -1$, we get $\nabla h_1(\mathbf{x}) = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$ and $\nabla h_2(\mathbf{x}) = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$

In both cases, two vectors $\nabla h_1(\mathbf{x})$ and $\nabla h_2(\mathbf{x})$ are not multiply of each other. Hence, the set $J(\mathbf{x}) = \{\nabla h_1(\mathbf{x}), \nabla h_2(\mathbf{x})\}$ is linearly independent. We conclude x is regular.

From all cases, we see that x is regular. Hence, all feasible points are regular. QED.

(b) KKT Conditions :

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu_1 \begin{pmatrix} -2x_1 - 2 \\ -2x_2 \end{pmatrix} + \mu_2 \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} = 0$$
$$\mu_1, \mu_2 \geq 0$$
$$\mu_i h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2$$

(c) We solve our KKT condition.

Case 1 None of h_1, h_2 are active at $\mathbf{x}$.

Then we have $\mu_1 = \mu_2 = 0$, so our KKT conditions become $\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$, a contradiction.

Case 2 Only h_1 that is active at $\mathbf{x}$.

Then we have $\mu_2 = 0$, so our KKT conditions become $\mu_1 \begin{pmatrix} 2x_1 + 2 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Together with the fact that h_1 is active, we have altogether three equations :

$$2x_1 + 2 = \frac{1}{\mu_1} \tag{1}$$
$$2x_2 = \frac{1}{\mu_1} \tag{2}$$
$$(x_1 + 1)^2 + x_2^2 = 1 \tag{3}$$

From third equation, we have $(2x_1 + 2)^2 + (2x_2)^2 = 4$, so we have $2\left(\frac{1}{\mu_1}\right)^2 = 4$. Since $\mu_1 \geq 0$, we see that $\mu_1 = \frac{1}{\sqrt{2}}$. Hence, $x_1 = \frac{\sqrt{2}-2}{2}$ and $x_2 = \frac{\sqrt{2}}{2}$. Since $x_1^2 + 2x_2^2 = \frac{3}{2} - \sqrt{2} + 1 < 3$ it follows that $\mathbf{x} = \begin{pmatrix} \frac{\sqrt{2}-2}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$ is indeed a feasible point. Hence, $\mathbf{x}$ is a KKT point.

Case 3 Only h_2 is active at $\mathbf{x}$.

Then we have $\mu_1 = 0$, so our KKT conditions become $\mu_2 \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$. Together with the fact that h_2 is active, we have altogether three equations :

$$2x_1 = \frac{-1}{\mu_2} \tag{1}$$
$$4x_2 = \frac{-1}{\mu_2} \tag{2}$$
$$x_1^2 + 2x_2^2 = 3 \tag{3}$$

Then we have $x_1 = \frac{-1}{2\mu_2}$ and $x_2 = \frac{-1}{4\mu_2}$ so substitute to third equation, we have $\left(\frac{1}{4} + \frac{1}{8}\right)\frac{1}{\mu_2^2} = 3$. Since $\mu_2 \geq 0$ then we have $\mu_2 = \frac{1}{\sqrt{8}}$. From there, we get $x_1 = -\sqrt{2}$ and $x_2 = -\frac{\sqrt{2}}{2}$. Since $(x_1 + 1)^2 + x_2^2 = \frac{7}{2} - 2\sqrt{2} < 1$ it follows that the point that we get is not a feasible point.

Case 4 h_1, h_2 are active at $\mathbf{x}$.

From the statement (3) at part a, at case 4, we see that if h_1, h_2 are active, then $x_1 = 1$ and $x_2 = \pm 1$. Furthermore, from out KKT conditions we have

$$\mu_1(2x_1 + 2) - \mu_2(2x_1) = 1 \tag{1}$$
$$\mu_1(2x_2) - \mu_2(4x_2) = 1 \tag{2}$$

If $x_1 = 1$ and $x_2 = 1$ we have $4\mu_1 - 2\mu_2 = 1$ and $2\mu_1 - 4\mu_2 = 1$. Solving two equations, we get $\mu_1 = \mu_2 = -\frac{1}{2}$ which is a contradiction since $\mu_1, \mu_2 \geq 0$.
If $x_1 = 1$ and $x_2 = -1$ we have $4\mu_1 - 2\mu_2 = 1$ and $-2\mu_1 + 4\mu_2 = 1$ which in turn give us $\mu_1 = \mu_2 = \frac{1}{2}$.

Altogether, there are two KKT points, namely $(\mathbf{x}; \mu) = \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}; \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\right), \left(\begin{pmatrix} \frac{\sqrt{2}-2}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}; \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}\right)$

(d) • $(\mathbf{x}; \mu) = \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}; \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\right)$

We have $H_L(\mathbf{x}) = \mu_1 \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} + \mu_2 \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}$. Now, $H_L(\mathbf{x})$ is a positive semidefinite matrix. Hence, the second order condition is satisfied; i.e $\mathbf{y}^T H_L(\mathbf{x})\mathbf{y} \geq 0$ for all $y \in T(\mathbf{x})$ where $T(\mathbf{x})$ is the tangent space to the feasible set at $\mathbf{x}$

- $(\mathbf{x};\mu) = \left(\begin{pmatrix} \frac{\sqrt{2}-2}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}; \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \right)$
  We have $H_L(\mathbf{x}) = \begin{pmatrix} -\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}$ which is negative semidefinite. In other words, $\mathbf{y}^T H_L(\mathbf{x})\mathbf{y} \geq 0$ for all $y \in T(\mathbf{x})$ where $T(\mathbf{x})$ is the tangent space to the feasible set at $\mathbf{x}$ if and only if $\mathbf{y}^T H_L(\mathbf{x})\mathbf{y} = 0$ for all $y \in T(\mathbf{x})$. Now, since h_2 is inactive at $\mathbf{x}$, we have $T(\mathbf{x}) = \{\mathbf{y} \mid \nabla h_1(\mathbf{x})^T y = 0\}$. Since $\nabla h_1(\mathbf{x}) = \begin{pmatrix} -\sqrt{2} \\ -\sqrt{2} \end{pmatrix}$, it follows that $\left\{ \begin{pmatrix} u \\ -u \end{pmatrix} \mid u \in \mathbb{R} \right\} = T(\mathbf{x})$. Take any vector from $T(\mathbf{x})$, say $\begin{pmatrix} y \\ -y \end{pmatrix}$. Then, we have $\begin{pmatrix} y & -y \end{pmatrix} \begin{pmatrix} -\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} y \\ -y \end{pmatrix} = \mathbf{0}$. Hence, $H_L(\mathbf{x})$ is positive semidefinite on $T(\mathbf{x})$ and the second order KKT condition is satisfied.

Question 6

(a) The Lagrangian function is $L(\mathbf{x},\mu) = -3x_1 - 2x_2 + \mu(x_1^2 + x_1x_2 + x_2^2 - 4)$ for $\mu \geq 0$. We shall minimize the function L on $x = (x_1; x_2)$ over the set X. Now if $\mu = 0$, it follows that $L(\mu) = -3x_1 - 2x_2$. Since taking $x_2 = 0$ and taking $x_1 \to \infty$, leads us to $L(\mathbf{x},\mu) = -\infty$, it follows that $inf\{L(\mathbf{x},\mu)\} = -\infty$. If $\mu > 0$, it follows that $L(\mathbf{x},\mu) = \mathbf{x}^T \begin{pmatrix} \mu & \frac{\mu}{2} \\ \frac{\mu}{2} & \mu \end{pmatrix} \mathbf{x} - \begin{pmatrix} 3 & 2 \end{pmatrix} \mathbf{x} - 4\mu$. Hence, $L(\mathbf{x},\mu)$ is a quadratic function. Furthermore, the matrix $\begin{pmatrix} 2\mu & \mu \\ \mu & 2\mu \end{pmatrix}$ has $\Delta_1 = 2\mu$ and $\Delta_2 = 3\mu^2$ which both are positive. Hence, $L(\mathbf{x},\mu)$ has a global minimizer over $\mathbb{R}^2$ and it occurs when $\begin{pmatrix} 2\mu & \mu \\ \mu & 2\mu \end{pmatrix} \mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$. We have a solution $\mathbf{x} = \begin{pmatrix} \frac{4}{3\mu} \\ \frac{1}{3\mu} \end{pmatrix}$ which is inside X. Hence, the global minimizer of L over $\mathbb{R}^2$ is also inside X. Therefore, the infimum is achieved when $\mathbf{x}$ is a global minimizer of L over $\mathbb{R}^2$.
Now we have $L(\mathbf{x},\mu) = \mathbf{x}^T \begin{pmatrix} \mu & \frac{\mu}{2} \\ \frac{\mu}{2} & \mu \end{pmatrix} \mathbf{x} - \begin{pmatrix} 3 & 2 \end{pmatrix} \mathbf{x} - 4\mu = \frac{1}{2}\mathbf{x}^T \begin{pmatrix} 3 \\ 2 \end{pmatrix} - \begin{pmatrix} 3 & 2 \end{pmatrix} \mathbf{x} - 4\mu = -\begin{pmatrix} \frac{3}{2} & 1 \end{pmatrix} \mathbf{x} - 4\mu = -\frac{7}{3\mu} - 4\mu$.
Hence, the Lagrangian Dual Problem is given by :

$$\begin{array}{rl} \max & \theta(\mu) = -\frac{7}{3\mu} - 4\mu \\ \text{s.t} & \mu > 0 \end{array}$$

(b) This has a lot of ways to do that. We can use Calculus, that is the first derivative must be zero. One can also make a no-square is negative equation. The findings presented here uses the AM-GM inequality.
By the AM-GM inequality, we get $-\theta(\mu) = \frac{7}{3\mu} + 4\mu \geq 2\sqrt{\frac{28}{3}}$ with equality happens when $(\mu^*)^2 = \frac{7}{12}$ which leads to $\mu^* = \sqrt{\frac{7}{12}}$. Hence, $\theta(\mu) = -\frac{7}{3\mu} - 4\mu$ is maximized when $\mu^* = \sqrt{\frac{7}{12}}$.

(c) Let $\mathbf{x}^* = \begin{pmatrix} \frac{4}{3\mu^*} \\ \frac{1}{3\mu^*} \end{pmatrix}$. Since $f(\mathbf{x}^*) = -\frac{14}{3\mu^*} = -2\sqrt{\frac{28}{3}}$, it follows that $f(\mathbf{x}^*) = \theta(\mu^*)$. Hence, the $\mathbf{x}^*$ is a global solution to the NLP.

Question 7

(a) Since $\mathbf{x}^*$ is a KKT point, we have $\mu^* h(\mathbf{x}^*) = 0$. Furthermore, we also have $g(\mathbf{x}^*) = 0$. Hence $L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*) + \lambda^* g(\mathbf{x}^*) + \mu^* h(\mathbf{x}^*) = f(\mathbf{x}^*)$. QED.

(b) Since $\mathbf{x}^*$ is feasible, we have $\mu h(\mathbf{x}^*) \leq 0$ and $g(\mathbf{x}^*) = 0$. Hence, we have $L(\mathbf{x}^*, \lambda, \mu) = f(\mathbf{x}^*) + \lambda g(\mathbf{x}^*) + \mu h(\mathbf{x}^*) \leq f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*)$. The last equality comes from part a. QED.

(c) Claim : $L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*)$.
Proof of claim : since f, h is convex and g is affine, we have :

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)\nabla f(\mathbf{x}^*) \\ \lambda^* g(\mathbf{x}) &= \lambda^* g(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)\lambda^* \nabla g(\mathbf{x}^*) \\ \mu^* h(\mathbf{x}) &\geq \mu^* h(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)\mu^* \nabla h(\mathbf{x}^*) \end{aligned}$$

Taking the sum of three equations, we get $L(\mathbf{x}, \lambda^*, \mu^*) \geq L(\mathbf{x}^*, \lambda^*, \mu^*) + (\mathbf{x} - \mathbf{x}^*)(\nabla f(\mathbf{x}^*) + \lambda^* \nabla g(\mathbf{x}^*) + \mu^* \nabla h(\mathbf{x}^*)) = L(\mathbf{x}^*, \lambda^*, \mu^*)$
Combining that with inequality at part b, we get the result. QED.