NATIONAL UNIVERSITY OF SINGAPORE
MATHEMATICS SOCIETY

PAST YEAR PAPER SOLUTIONS
with credits to Sean Lim Wei Xinq, Zheng Shaoxuan

MA3236 Non-linear Programming
AY 2009/2010 Sem 1

Throughout this paper, for any matrix A, $A > 0$ denotes A is positive definite, $A \geq 0$ denotes A is positive semidefinite, $A < 0$ denotes A is negative definite, $A \leq 0$ denotes A is negative semidefinite and $A \approx 0$ denotes A is indefinite.

Question 1

(a) We have

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix}$$

Then, setting $\nabla f(\mathbf{x}) = 0$ gives us

$$2x_1 - x_2 + 3x_1^2 + 2x_1^3 = 0 \tag{1}$$
$$x_1 = x_2 \tag{2}$$

From (1) we have

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

So the stationary points are $\mathbf{x}_*^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\mathbf{x}_*^{(2)} = \begin{pmatrix} -1/2 \\ -1/2 \end{pmatrix}$ and $\mathbf{x}_*^{(3)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$.

(b) The Hessian of f is given by

$$H_f(\mathbf{x}) = \begin{pmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{pmatrix}$$

So we have

$$\begin{aligned} H_f(\mathbf{x}_*^{(1)}) &= \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix} \\ H_f(\mathbf{x}_*^{(2)}) &= \begin{pmatrix} 1 & -2 \\ -2 & 2 \end{pmatrix} \\ H_f(\mathbf{x}_*^{(3)}) &= \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix} \end{aligned}$$

For $H_f(\mathbf{x}_*^{(1)})$ and $H_f(\mathbf{x}_*^{(3)})$, we have $\Delta_1 = \Delta_2 = 4 > 0$, and hence, $H_f(\mathbf{x}_*^{(1)}) = H_f(\mathbf{x}_*^{(3)}) > 0$, which implies that $\mathbf{x}_*^{(1)}$ and $\mathbf{x}_*^{(3)}$ are strict local minimizers. Then, $\det(H_f(\mathbf{x}_*^{(2)})) < 0$ implies $H_f(\mathbf{x}_*^{(2)}) \approx 0$, which implies that $\mathbf{x}_*^{(2)}$ is a saddle point.

(c) We can write f as

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + x_1^4 + 2x_1^3$$

where

$$Q = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Then, note that $\mathbf{Q}$ is symmetric positive definite. So if $\lambda_{min}(\mathbf{Q}) > 0$ is the smallest eigenvalue of $\mathbf{Q}$, we have

$$f(\mathbf{x}) \geq \lambda_{min}(\mathbf{Q})||\mathbf{x}||^2 + x_1^4 + 2x_1^3$$

From here, we can observe that f is coercive. Hence, f has a global minimizer but no global maximizer.

Question 2

Since f is continuous and coercive, it follows that for all $K > 0$, we can find a $k > 0$ such that if $||\mathbf{x}|| > k$, then $f(\mathbf{x}) > K$. Let $K > 0$ and $x_0 \in D$ such that $K > f(x_0)$. Now consider the set

$$S := \{\mathbf{x} \in D \mid ||\mathbf{x}|| \leq k\}$$

Note that S is non-empty and bounded. Also, S is closed by definition of S. Hence, by Weierstrass Theorem, there is a global minimizer in S, i.e. there is an $\mathbf{x}^* \in S$ such that

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in S$$

Now if $\mathbf{x} \notin S$, then we have

$$f(\mathbf{x}^*) \leq K < f(\mathbf{x})$$

which shows that $\mathbf{x}^*$ is also a global minimizer in D.

Question 3

(a) For any $\mathbf{x} \in \mathbb{R}^q$ we have, for any $\mathbf{B} \in \mathbb{R}^{n \times q}$ we have

$$\mathbf{x}^T \mathbf{B}^T \mathbf{A} \mathbf{B} \mathbf{x} = (\mathbf{B}\mathbf{x})^T \mathbf{A} (\mathbf{B}\mathbf{x}) \geq 0$$

since A is positive definite. This shows that $\mathbf{B}^T\mathbf{A}\mathbf{B}$ is positive semidefinite.

To have the fact that $\mathbf{B}^T\mathbf{A}\mathbf{B} > 0$, we require that $\mathbf{B}\mathbf{x} \neq \mathbf{0}$ whenever $\mathbf{x} \neq \mathbf{0}$. This will happen if $\mathcal{N}(\mathbf{B}) = \{\mathbf{0}\}$, where $\mathcal{N}$ denotes the nullspace of $\mathbf{B}$. Hence, $\mathbf{A}_k$ must be positive definite since $\mathbf{A}_k = \mathbf{I}_{n\times k}^T \mathbf{A} \mathbf{I}_{n \times k}$, where $\mathbf{I}_{n\times k} = (b_{ij})$ where

$$b_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and $\mathcal{N}(\mathbf{I}_{n\times k}) = \{\mathbf{0}\}$ for all $k = 1, \cdots, n$.

(b) Let $\mathbf{P}$ be an orthogonal matrix such that $\mathbf{A} = \mathbf{P}^T\mathbf{D}\mathbf{P}$ for some diagonal matrix D with its diagonal entries $\lambda_1, \cdots, \lambda_n$ as the eigenvalues of $\mathbf{A}$ and $\mathbf{P}_k = \mathbf{P}\mathbf{I}_{n\times k}$ where $\mathbf{I}_{n\times k}$ is defined as in part (a) so that $\mathbf{A}_k = \mathbf{P}_k^T\mathbf{D}\mathbf{P}_k$. If μ is an eigenvalue of $\mathbf{A}_k$ and $\mathbf{v}$ its corresponding eigenvector, then we have

$$\begin{aligned}
\mu\mathbf{v} &= \mathbf{A}_k\mathbf{v}\\
\mathbf{v}^T\mu\mathbf{v} &= \mathbf{v}^T\mathbf{A}_k\mathbf{v}\\
&= \mathbf{v}^T\mathbf{P}_k^T\mathbf{D}\mathbf{P}_k\mathbf{v}\\
&= \mathbf{x}^T\mathbf{D}\mathbf{x}
\end{aligned}$$

for some vector $\mathbf{x} \in \mathbb{R}^n$. Then we have

$$\begin{aligned}
\mu||\mathbf{v}||^2 &= \sum_{i=1}^n \lambda_i\mathbf{x}_i^2\\
&\leq \sum_{i=1}^n \lambda_{max}\mathbf{x}_i^2 = \lambda_{max}||x||^2 = \lambda_{max}||v||^2
\end{aligned}$$

Hence we have $\mu \leq \lambda_{max}$. Similarly, we have

$$\begin{aligned}
\mu||\mathbf{v}||^2 &= \sum_{i=1}^n \lambda_i\mathbf{x}_i^2\\
&\geq \sum_{i=1}^n \lambda_{min}\mathbf{x}_i^2 = \lambda_{min}||x||^2 = \lambda_{min}||v||^2
\end{aligned}$$

So $\mu \geq \lambda_{min}$ and this completes the proof.

(c) By Taylor's Theorem, there exists $\mathbf{w} \in [\mathbf{x}, \mathbf{x}_k]$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k\rangle + \frac{1}{2}\langle H_f(\mathbf{w}), \mathbf{x} - \mathbf{x}_k\rangle$$

for all $k = 1, \cdots, p$. Since f is convex, $H_f(\mathbf{x}) \geq 0$ and hence, $\langle H_f(\mathbf{w}), \mathbf{x} - \mathbf{x}_k\rangle \geq 0$ and therefore

$$f(\mathbf{x}) \geq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k\rangle \quad \forall \mathbf{x} \in S \quad \forall 1 \leq k \leq p$$

Hence, we have

$$f(\mathbf{x}) \geq \max\{L_k(\mathbf{x}) := f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k\rangle \mid 1 \leq k \leq p\} \quad \forall \mathbf{x} \in S$$

Question 4

(a) (Correction in paper during day of exam: "Width is twice the length" instead of "Length is twice the width")

Let x_1, x_2 and x_3 denote the length, width and height of the box respectively. Now since the width is twice the length, we have $2x_1 = x_2$. Then the volume of the box, $V(\mathbf{x})$ is given by

$$V(\mathbf{x}) = x_1x_2x_3 = 2x_1^2x_3$$

Since it is not necessary to use all the wire and paper, we have inequality constraints instead of equality. For the wire, we have

$$4x_1 + 4x_2 + 4x_3 - 20 = 4x_1 + 8x_1 + 4x_3 - 20 = 12x_1 + 4x_3 - 20 \leq 0$$

Dividing by 4 gives us

$$h_1(\mathbf{x}) := 3x_1 + x_3 - 5 \leq 0$$

Next we consider the paper, which is the total surface area. We have

$$2x_1x_2 + 2x_1x_3 + 2x_2x_3 - 16 = 2x_1(2x_1) + 2x_1x_3 + 2(2x_1)x_3 - 16 = 4x_1^2 + 6x_1x_3 - 16 \leq 0$$

Dividing by 2 gives us

$$h_2(\mathbf{x}) := 2x_1^2 + 3x_1x_3 - 8 \leq 0$$

Next, since x_1 and x_3 represent the length and height of the box respectively, we require that they are non-negative, i.e. $x_1, x_3 \geq 0$.

The objective function, if we are translating the problem into a minimization problem is the negation of the volume, i.e.

$$f(\mathbf{x}) := -V(\mathbf{x}) = -2x_1^2x_3, \quad x \in \mathbb{R}^2$$

Hence our optimization problem is

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & h_1(\mathbf{x}) \leq 0 \\ & h_2(\mathbf{x}) \leq 0 \qquad\qquad\qquad\qquad x_1, x_3 \geq 0 \end{array}$$

(b) The feasible set S is given by

$$S := \{\mathbf{x} \in \mathbb{R}^2 \mid h_1(\mathbf{x}) \leq 0,\ h_2(\mathbf{x}) \leq 0, x_1, x_3 \geq 0\}$$

To show that all feasible points are regular, we consider 4 cases. Define the set $R(\mathbf{x})$ by

$$R(\mathbf{x}) = \{\nabla h_i(\mathbf{x})\}$$

where $h_i(\mathbf{x}) = 0, i = 1, 2$. Now we have

$$\nabla h_1(\mathbf{x}) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$
$$\nabla h_2(\mathbf{x}) = \begin{pmatrix} 4x_1 + 3x_3 \\ 3x_1 \end{pmatrix}$$

Case 1: $J(\mathbf{x}) = \emptyset$. Then $R(\mathbf{x}) = \emptyset$ which is linearly independent by definition.

Case 2: $J(\mathbf{x}) = \{1\}$. Then $R(\mathbf{x}) = \{\nabla h_1(\mathbf{x})\}$, and $\nabla h_1(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^2$. So the set is linearly independent.

Case 3: $J(\mathbf{x}) = \{2\}$. Then $R(\mathbf{x}) = \{\nabla h_2(\mathbf{x})\}$, and $\nabla h_2(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$. But $J(\mathbf{x}) = \{2\}$ implies that $h_2(\mathbf{x}) = 0 \Rightarrow -8 = 0$, a contradiction. So the set is always linearly independent.

Case 4: $J(\mathbf{x}) = \{1, 2\}$. Then $R(\mathbf{x}) = \{\nabla h_1(\mathbf{x}), \nabla h_2(\mathbf{x})\}$. Now consider the matrix

$$\mathcal{R}(\mathbf{x}) := \begin{pmatrix} 3 & 4x_1 + 3x_3 \\ 1 & 3x_1 \end{pmatrix}$$

Then, the set $R(\mathbf{x})$ is linearly independent if and only if the matrix $\mathcal{R}(\mathbf{x})$ is invertible. Now $\det(\mathcal{R}(\mathbf{x})) = 5x_1 - 3x_3 = 0$ if and only if $x_1 = \frac{3}{5}x_3$. But this implies

$$\begin{aligned} h_1(\mathbf{x}) = 3\left(\frac{3}{5}x_3\right) + x_3 - 5 &= 0 \\ \frac{14}{5}x_3 &= 5 \\ x_3 &= \frac{25}{14} \end{aligned}$$

Then, we have $x_1 = \frac{15}{14}$. But this gives us

$$h_2(\mathbf{x}) = 2\left(\frac{15}{14}\right)^2 + 3\left(\frac{15}{14}\right)\left(\frac{25}{14}\right) - 8 \neq 0$$

which cannot happen since $J(\mathbf{x}) = \{1, 2\}$.

So in all cases, we have shown that $\mathbf{x} \in S$ must be regular.

(c) We have $\nabla f(\mathbf{x}) = \begin{pmatrix} -4x_1x_3 \\ -2x_1^2 \end{pmatrix}$. Hence the KKT conditions are

$$\begin{pmatrix} -4x_1x_3 \\ -2x_1^2 \end{pmatrix} + \mu_1 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 4x_1 + 3x_3 \\ 3x_1 \end{pmatrix} = \mathbf{0}, \quad \mu_1, \mu_2 \geq 0$$

Case 1: $J(\mathbf{x}) = \emptyset$. Then $\mu_1 = \mu_2 = 0$ and we have

$$\begin{aligned} -4x_1x_3 &= 0 \\ -2x_1^2 &= 0 \end{aligned}$$

But this implies $x_1 = x_3 = 0$, which contradicts our assumption that $x_1, x_3 > 0$.

Case 2: $J(\mathbf{x}) = \{1\}$. Then $\mu_2 = 0$ and $\mu_1 > 0$ and we have

$$\begin{aligned} -4x_1x_3 + 3\mu_1 &= 0 && (3) \\ -2x_1^2 + \mu_1 &= 0 && (4) \\ 3x_1 + x_3 - 5 &= 0 && (5) \end{aligned}$$

Now (4) implies $\mu_1 = 2x_1^2$. Then, from (3) we have $-4x_1x_3 + 6x_1^2 = 0 \Rightarrow x_1(-4x_3 + 6x_1) = 0 \Rightarrow x_1 = 0$ or $3x_1 = 2x_3$.

If $x_1 = 0$, then from (3) we have $\mu_1 = 0$, a contradiction!

If $3x_1 = 2x_3$, then from (5) we have $x_3 = 5/3$, which then gives us $x_1 = 10/9$. Then, $\mu_1 = 200/81$. But observe that

$$h_2\left(\begin{pmatrix} 10/9 \\ 5/3 \end{pmatrix}\right) = 2\left(\frac{10}{9}\right)^2 + 3\left(\frac{10}{9}\right)\left(\frac{5}{3}\right) - 8 = \frac{2}{81} \not\leq 0$$

Hence it is NOT feasible.

Case 3: $J(\mathbf{x}) = \{2\}$. Then $\mu_1 = 0$ and $\mu_2 > 0$ and we have

$$-4x_1x_3 + 4\mu_2 x_1 + 3\mu_2 x_3 = 0 \tag{6}$$
$$-2x_1^2 + 3x_1\mu_2 = 0 \tag{7}$$
$$2x_1^2 + 3x_1x_3 - 8 = 0 \tag{8}$$

Here, (7) implies $x_1(-2x_1 + 3\mu_2) = 0$ which implies that either $x_1 = 0$ or $3\mu_2 = 2x_1$. But $x_1 = 0$ implies $-8 = 0$, a contradiction.
But $3\mu_2 = 2x_1$ implies from (6) that

$$-2x_1x_3 + \frac{8}{3}x_1^2 = 0 \tag{9}$$

Then, 2(8) + 3(9) implies $12x_1^2 = 16 \Rightarrow x_1^2 = 4/3 \Rightarrow x_1 = 2/\sqrt{3}$. This gives us $\mu_2 = 4/3\sqrt{3}$ and from (8) we have

$$\begin{aligned} \frac{8}{3} + 4\sqrt{3}x_3 - 8 &= 0 \\ x_3 &= \frac{16}{3}\frac{1}{4\sqrt{3}} = \frac{4}{3\sqrt{3}} \end{aligned}$$

Case 4: $J(x) = \{1, 2\}$. Then $\mu_1, \mu_2 > 0$ and we have

$$-4x_1x_3 + 3\mu_1 + 4\mu_2 x_1 + 3\mu_2 x_3 = 0 \tag{10}$$
$$-2x_1^2 + \mu_1 + 3\mu_2 x_1 = 0 \tag{11}$$
$$3x_1 + x_3 - 5 = 0 \tag{12}$$
$$2x_1^2 + 3x_1x_3 - 8 = 0 \tag{13}$$

Now from (12) we have $x_3 = 5 - 3x_1$, and from (13) we have

$$\begin{aligned} 2x_1^2 + 3x_1(5 - 3x_1) - 8 &= 0 \\ -7x_1^2 + 15x_1 - 8 &= 0 \\ 7x_1^2 - 15x_1 + 8 &= 0 \\ (7x_1 - 8)(x_1 - 1) &= 0 \\ x_1 &= \frac{8}{7}, 1 \end{aligned}$$

If $x_1 = 1$ then $x_3 = 2$ and we have from (10) and (11)

$$\begin{aligned} 3\mu_1 + 10\mu_2 &= 8 \\ \mu_1 + 3\mu_2 &= 2 \end{aligned}$$

Solving the simultaneous equation yields $\mu_1 = -4, \mu_2 = 2$, a contradiction since $\mu_1 > 0$.
If $x_1 = 8/7$, then $x_3 = 11/7$ and we have from (10) and (11)

$$\begin{aligned} 3\mu_1 + \frac{65}{7}\mu_2 &= \frac{352}{49} \\ \mu_1 + \frac{24}{7}\mu_2 &= \frac{128}{49} \end{aligned}$$

Solving the simultaneous equation yields $\mu_1 = 128/343, \mu_2 = 32/49$. Check again that $\begin{pmatrix} 8/7 \\ 11/7 \end{pmatrix}$ is feasible.

Hence the KKT points are $\mathbf{x}_1^* = \begin{pmatrix} 2/\sqrt{3} \\ 4/3\sqrt{3} \end{pmatrix}$ and $\mathbf{x}_2^* = \begin{pmatrix} 8/7 \\ 11/7 \end{pmatrix}$.

(d) Let the sets S_1 and S_2 be defined as

$$S_1 := \{x \in \mathbb{R}^2 \mid h_1(\mathbf{x}) \leq 0, x_1 \geq 0, x_3 \geq 0\}$$

$$S_2 := \{x \in \mathbb{R}^2 \mid h_2(\mathbf{x}) \leq 0\}$$

Now observe that S_1 is closed and bounded and S_2 is closed since h_2 is continuous, and that $S = S_1 \cap S_2$. Since S_1 is bounded, $S \subseteq S_1$ implies that S is also bounded, and S is closed since the intersection of closed sets is closed. So by Weierstrass Theorem, there is a global minimizer. Since all feasible points are regular, it follows that every global minimizer is a local minimizer, and by the KKT necessary conditions, every global min is a KKT point. Hence we simply evaluate

$$f(\mathbf{x}_1^*) = -\frac{32}{9\sqrt{3}} = -2.053$$
$$f(\mathbf{x}_2^*) = -\frac{1408}{343} = -4.105$$

So the global minimizer is $\mathbf{x}_2^*$.

Question 5

(a) Since it is a convex programming problem, a KKT point is also a global minimizer. Let $h_1(\mathbf{x}) := x_1^2 + 4x_2^2 - 4 + \epsilon$, $h_2(\mathbf{x}) = -x_1$ and $h_3(\mathbf{x}) = -x_2$. Now we have

$$\nabla f(\mathbf{x}) = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$$
$$\nabla h_1(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix}$$
$$\nabla h_2(\mathbf{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$
$$\nabla h_3(\mathbf{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

So the KKT conditions are

$$\begin{pmatrix} -2 \\ -4 \end{pmatrix} + \mu_1 \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \mathbf{0}$$

where $\mu_1, \mu_2, \mu_3 \geq 0$.

Case 1: $J(\mathbf{x}) = \emptyset$. Then $\mu_1 = \mu_2 = \mu_3 = 0$ and hence, $\begin{pmatrix} -2 \\ -4 \end{pmatrix} = \mathbf{0}$, a contradiction.

Case 2: $J(\mathbf{x}) = \{1\}$. Then $\mu_1 > 0$ and $\mu_2 = \mu_3 = 0$ and hence,

$$-2 + 2\mu_1 x_1 = 0 \tag{14}$$
$$-4 + 8\mu_1 x_2 = 0 \tag{15}$$
$$x_1^2 + 4x_2^2 - 4 + \epsilon = 0 \tag{16}$$
$$x_1, x_2 > 0 \tag{17}$$

Now (14) implies that $x_1 = 1/\mu_1$ and (15) implies that $x_2 = 1/2\mu_1$. Substituting them into (16) yields

$$\begin{aligned}
\frac{1}{\mu_1^2} + \frac{1}{\mu_1^2} - 4 + \epsilon = 0 \\
\frac{2}{\mu_1^2} &= 4 - \epsilon \\
\mu_1^2 &= \frac{2}{4-\epsilon} \\
\mu_1 &= \sqrt{\frac{2}{4-\epsilon}}
\end{aligned}$$

Then we have

$$x_1 = \sqrt{\frac{4-\epsilon}{2}} \quad \text{and} \quad x_2 = \frac{\sqrt{4-\epsilon}}{2\sqrt{2}}$$

From here, we have a KKT point and hence a global minimizer. If there is another KKT point, then their function values will be the same. So we have

$$x^*(\epsilon) = \begin{pmatrix} \sqrt{(4-\epsilon)/2} \\ \sqrt{4-\epsilon}/2\sqrt{2} \end{pmatrix}$$

with the multipliers $\mu_1^*(\epsilon) = \sqrt{2/(4-\epsilon)}$, $\mu_2^*(\epsilon) = \mu_3^*(\epsilon) = 0$.

(b) Let $F(\epsilon) = f(x^*(\epsilon))$. Then

$$F(\epsilon) = -2\sqrt{\frac{4-\epsilon}{2}} - 4\frac{\sqrt{4-\epsilon}}{2\sqrt{2}} = -4\sqrt{\frac{4-\epsilon}{2}}$$

Then,

$$F'(\epsilon) = \frac{-2\sqrt{2}}{\sqrt{4-\epsilon}} \cdot -\frac{1}{2}$$

by the chain rule. Simplifying gives us

$$F'(\epsilon) = \sqrt{\frac{2}{4-\epsilon}} = \mu_1^*(\epsilon)$$

as required.

Question 6

(a) We are given

$$f(\mathbf{x}) = 4x_1^2 + 4x_2^2 - 2x_1x_2 + x_3^2$$

Observe that

$$f(\mathbf{x}) = \mathbf{x}^T\left(\frac{1}{2}\mathbf{Q}\right)\mathbf{x} = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$$

where

$$\mathbf{Q} = \begin{pmatrix} 8 & -2 & 0 \\ -2 & 8 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

(You may want to check that $\mathbf{Q}$ is indeed symmetric positive definite)

(b) By Proposition 5.1(a) we have

$$\frac{E(\mathbf{x}_{k+1})}{E(\mathbf{x}_k)} = 1 - \frac{\langle \nabla q(\mathbf{x}_k), \nabla q(\mathbf{x}_k)\rangle^2}{\langle \nabla q(\mathbf{x}_k), \mathbf{Q}\nabla q(\mathbf{x}_k)\rangle \langle \nabla q(\mathbf{x}_k), \mathbf{Q}^{-1}\nabla q(\mathbf{x}_k)\rangle}$$

Since f is a quadratic function, then its quadratic Taylor expansion, q is equal to the function itself. Also, by definition of $E(\mathbf{x}_k)$ and $\mathbf{d}_k$ we have

$$\frac{f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)}{f(\mathbf{x}_k) - f(\mathbf{x}^*)} = 1 - \frac{||\mathbf{d}_k||^4}{\langle \mathbf{d}_k, \mathbf{Q}\mathbf{d}_k\rangle (\nabla q(\mathbf{x}_k)^T \mathbf{Q}^{-1} \nabla q(\mathbf{x}_k))}$$

Now $\nabla f(\mathbf{x}^*) = \mathbf{0}$ implies $\mathbf{x}^* = \mathbf{0}$, and hence, $f(\mathbf{x}^*) = 0$, and so we have

$$\frac{f(\mathbf{x}_{k+1})}{f(\mathbf{x}_k)} = 1 - \frac{||\mathbf{d}_k||^4}{\langle \mathbf{d}_k, \mathbf{Q}\mathbf{d}_k\rangle (\nabla q(\mathbf{x}_k)^T \mathbf{Q}^{-1} \nabla q(\mathbf{x}_k))}$$

Next, we multiply both sides by $f(\mathbf{x}_k)$ to obtain

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) - \frac{||\mathbf{d}_k||^4 (f(\mathbf{x}_k))}{\langle \mathbf{d}_k, \mathbf{Q}\mathbf{d}_k\rangle (\nabla q(\mathbf{x}_k)^T \mathbf{Q}^{-1} \nabla q(\mathbf{x}_k))}$$

Then, by definition of f and $\nabla q(\mathbf{x}_k)$ we have

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) - \frac{1}{2}\frac{||\mathbf{d}_k||^4 (\mathbf{x}_k^T \mathbf{Q}\mathbf{x}_k)}{\langle \mathbf{d}_k, \mathbf{Q}\mathbf{d}_k\rangle \left(\mathbf{x}_k^T \mathbf{Q}^T \mathbf{Q}^{-1} \mathbf{Q}\mathbf{x}_k\right)}$$

Since $\mathbf{Q}$ is symmetric, it follows that $\mathbf{Q}^T = \mathbf{Q}$ and canceling from numerator and denominator gives us

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) - \frac{1}{2}\frac{||\mathbf{d}_k||^4}{\langle \mathbf{d}_k, \mathbf{Q}\mathbf{d}_k\rangle}$$

as required.

(c) We first determine the eigenvalues of $\mathbf{Q}$. We have

$$\begin{aligned}
\begin{vmatrix} 8-\lambda & -2 & 0 \\ -2 & 8-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix} &= 0 \\
(2-\lambda)\begin{vmatrix} 8-\lambda & -2 \\ -2 & 8-\lambda \end{vmatrix} &= 0 \\
\lambda = 2 \quad \text{or} \quad (8-\lambda)^2 - 4 &= 0 \\
\lambda &= 2, 6, 10
\end{aligned}$$

Now since $\kappa(\mathbf{Q}) = \lambda_{max}(\mathbf{Q})/\lambda_{min}(\mathbf{Q}) = 10/2 = 5$ is large, the convergence rate, ρ is given by

$$\rho(\mathbf{Q}) = 1 - \frac{4}{\kappa(\mathbf{Q})} = \frac{1}{5}$$

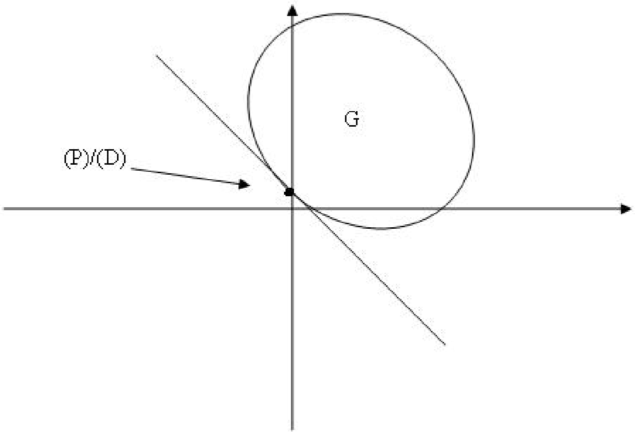
Here, hence, in the worst case, the number of iterations, k is given by

$$k = \left\lfloor \frac{\log \epsilon}{\log \rho(\mathbf{Q})} \right\rfloor + 1$$

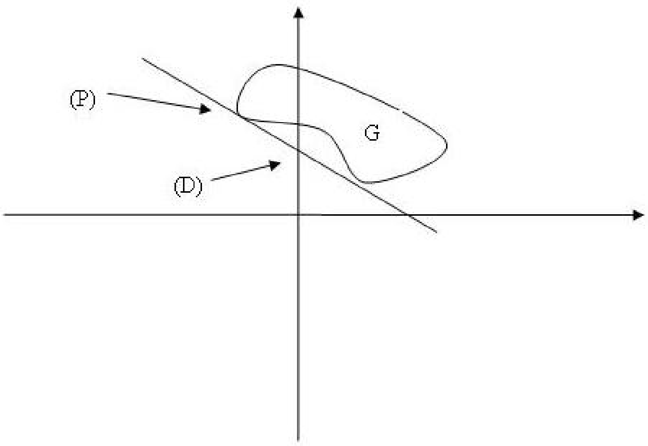
It is given that $\epsilon = 10^{-8}$, so $k = 12$.

Question 7

(a) (i) Consider the figure below.



(ii) Consider the figure below.



(b) The Lagrangian function, $L(\mathbf{x}, \mu)$ is given by

$$\begin{aligned} L(\mathbf{x}, \mu) &= -2x_1 - x_2 + \mu(x_1^2 + 4x_2^2 - 2x_1x_2 - 4) \\ &= \mu x_1^2 + 4\mu x_2^2 - 2\mu x_1 x_2 - 2x_1 - x_2 - 4\mu \end{aligned}$$

where $\mu \geq 0$ and $\mathbf{x} \in \mathbb{R}^2$.

Now we want to determine

$$\theta(\mu) = \inf_{x \in \mathbb{R}^2} L(\mathbf{x}, \mu)$$

Hence we have

$$L_{\mathbf{x}}(\mathbf{x}, \mu) = \begin{pmatrix} 2\mu x_1 - 2\mu x_2 - 2 \\ 8\mu x_2 - 2\mu x_1 - 1 \end{pmatrix}$$

Setting $L_{\mathbf{x}} = 0$ gives us

$$\mu x_1 - \mu x_2 - 1 = 0 \tag{18}$$
$$-2\mu x_1 + 8\mu x_2 - 1 = 0 \tag{19}$$

Taking (19) + 2(18) yields

$$6\mu x_2 - 3 = 0 \Rightarrow x_2 = \frac{1}{2\mu}$$

Note here that $\mu \neq 0$ since that would imply that $-1 = 0$, a contradiction! Then, we have $x_1 = 3/2\mu$. Hence we obtain

$$\begin{aligned}\theta(\mu) &= \frac{9}{4\mu} + \frac{1}{\mu} - \frac{3}{2\mu} - \frac{3}{\mu} - \frac{1}{2\mu} - 4\mu \\ &= -\frac{7}{4\mu} - 4\mu\end{aligned}$$

So the dual problem is

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

Then, we have

$$\theta'(\mu_*) = \frac{7}{4\mu_*^2} - 4 = 0 \Rightarrow \mu_*^2 = \frac{7}{16} \Rightarrow \mu_* = \frac{\sqrt{7}}{4}$$

(c) The Lagrangian function is given by

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

Now we want to find the infimum of L for $\mu \geq 0$. Observe that L is convex, hence, the minimizer is obtained at the stationary point. So we have

$$L_{\mathbf{x}} = \mathbf{Q}\mathbf{x} + \mathbf{c} - \mathbf{A}^T\lambda - \mu$$

Rewriting L gives us

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

Setting $L_{\mathbf{x}} = 0$ gives us

$$-\mathbf{Q}\mathbf{x} = \mathbf{c} - \mathbf{A}^T\lambda - \mu$$

and hence we have

$$\begin{aligned}\theta(\lambda, \mu) &= \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu) \\ &= \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\lambda - \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ &= -\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\lambda\end{aligned}$$

So the dual problem is

$$\begin{aligned}\max \quad & -\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\lambda \\ \text{s.t.} \quad & \mathbf{Q}\mathbf{x} + \mathbf{c} - \mathbf{A}^T\lambda - \mu = 0 \\ & \mu \geq 0\end{aligned}$$

as desired.